LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 200
aprile 1985
anno XVIII
volume XXXIV

Civiltà e follia

Criminologi, medici e freniatri italiani della fine del XIX secolo, interrogandosi sul rapporto esistente tra vita sociale e follia, pongono le fondamenta teoriche, professionali nonché istituzionali della psichiatria

di Metello Corulli

Il dibattito sulla civiltà nasce all'interno dello scenario sociale e politico post-unitario, in un momento storico nel quale le condizioni dello Stato italiano sono poco conosciute agli stessi governanti: per questo motivo si tenta di inventariare il patrimonio nazionale in tutti i campi e il Parlamento promuove una serie di inchieste per conoscere le condizioni dei possedimenti agrari, delle risorse naturali e commerciali, dei beni culturali: una grande mobilitazione progettuale di tutte le risorse intellettuali del periodo. L'ideologia del progresso, il tentativo di progettare sull'esistenza e di fondare una cultura laica trovano manifestazione storica nelle prime prometeiche esposizioni nazionali, come quella di Firenze del 1861.

I nuovi quadri tecnici della nazione partecipano così alla lotta tra oscurantismo e progresso, allo scenario del «Gran Ballo Excelsior».

I medici avevano avuto continue occasioni di scoperta delle realtà locali, dei problemi igienico-sanitari e della diversificazione regionale con le guerre d'indipendenza, le campagne per la repressione del brigantaggio, la mobilitazione per le gravi epidemie di malattie infettive e, successivamente, con l'organizzazione dell'esercito nazionale e le visite di leva. Da tutte queste osservazioni ha avuto origine il movimento igienista.

I freniatri, vinta almeno in parte l'ostilità degli ordini religiosi, dei colleghi clinici, dei giuristi e dei magistrati, intendono partecipare attivamente alla costituzione dello Stato post-unitario.

Accade così che tra la fine dell'Ottocento e i primissimi anni del Novecento criminologi, freniatri e medici italiani affrontano una tematica inquietante: «Veramente la civiltà scopre le piaghe del vivere sociale, non le occasiona, le scema, non le moltiplica; la civiltà medica, non tortura, risana, non uccide»?

Il termine «civiltà» comparso all'interno del clima culturale dell'Illuminismo francese a indicare la superiorità dell'Europa sul resto del mondo, la differenza nello spazio tra le varie culture, veniva ora utilizzato per la prima volta per proiettare l'ineguaglianza nel tempo, per indicare la superiorità del presente sul passato, il cammino verso uno stato di perfezione.

La scienza freniatrica italiana in fine secolo

Al quarto congresso della Società freniatrica italiana tenutosi a Voghera nel 1883, Andrea Verga nella relazione introduttiva esordisce: «Se quasi tutti i vecchi manicomi d'Italia sono oggidì sia nella parte materiale che nella morale rimodernati secondo i dettami della scienza e della civiltà... Il mio sogno che la redenzione nazionale avesse ad estendersi anche ai poveri pazzi, è per vostra virtù ampiamente realizzato... Sì, noi possiamo vantarci d'aver contribuito alla più santa delle riparazioni, d'aver assistito ad una resurrezione dei morti... Ormai non v'ha disciplina sulla quale la psicologia non stenda le sue grandi ali. La nostra scienza si confonde con quella dell'umanità.»

Si può comprendere l'entusiasmo di uno dei padri fondatori della psichiatria italiana, di Andrea Verga; la Società, fondata a Roma nell'ottobre 1873 all'interno dell'undicesimo e ultimo congresso degli scienziati italiani, annovera 109 membri effettivi, 12 membri onorari stranieri. Verga è portavoce dell'entusiasmo di una piccola comunità di scienziati e tuttavia abbastanza forte da avere ora in mano la gestione di tutti i manicomi d'Italia, da essersi guadagnata uno spazio sociale e professionale.

È una disciplina giovane: le comunicazioni vertono su singoli temi (sulla «mania transitoria»; sulla «temperatura dei pazzi»; sulle «circonvoluzioni e scissure verticali in un cervello di lipemaniaco»; sulla «trabecola cinerea nei pazzi»; sulla «cellula nervosa motrice», relazione di Camillo Golgi; ecc.). Manca ancora di tanto, se non altro di un metodo clinico e di una classificazione delle malattie mentali; eppure è una disciplina ambiziosa («La nostra scienza si confonde con quella dell'umanità»).

La relazione più consistente, più importante, più discussa, al centro del congresso è quella svolta dal direttore del

manicomio di Fermo, Roberto Adriani: «La educazione come profilassi della pazzia».

Il relatore appare insoddisfatto dei risultati ottenuti: l'incremento della prosperità pubblica, i progressi della scienza, la previdenza delle civili istituzioni e l'aumento dei mezzi di difesa contro i mali della vita avrebbero dovuto portare una progressiva diminuzione del numero dei pazzi, ma tale riduzione non sembra essersi verificata.

I dati statistici in possesso dei freniatri erano pochi e frammentari. Alcuni manicomi stampavano bollettini che sembravano documentare un incremento continuo del numero dei degenti; lo stesso fenomeno sembrava configurarsi nei primi censimenti effettuati dall'Archivio italiano per le malattie nervose e nelle ricerche, seppure incomplete, promosse da Verga; ma era difficile comprendere se l'incremento numerico non era semplicemente dovuto a un più ampio e accurato rilevamento statistico. Per altro sono questi i primi anni che la statistica da tecnica dell'amministrazione viene applicata alle scienze dell'uomo, diviene tecnica morale, espressione del bisogno di ordine e di efficienza.

Tuttavia Adriani avanza una interpretazione di quello che ai suoi occhi appare un incremento del numero dei folli: «La società moderna ha recato una maggiore somma di beni, ma ha aperto ad un tempo l'adito ad emozioni più vive e più frequenti ed a passioni più complesse; e così nelle industrie e nei commerci, nelle scienze e nelle arti, come nelle politiche e sociali controversie, è chiamata ad una attività mentale eccessiva, alla quale si abbandonano con febbrile ardore, in mezzo ai contrasti fra le speranze e i timori, le ansie, i disinganni e le angoscie che l'accompagnano.»

L'organo mentale può adattarsi alla maggiore attività e trasmettere alla prole i maggiori poteri della mente conseguiti oppure, inetto a sostenere il dispendio di forza che la maggiore attività intellettuale e morale richiederebbe, decade e infralisce.

«Succede così che mentre la generalità degli uomini cammina ad un progresso lento, impercettibile da una generazione all'altra, ma continuo e costante, vi ha d'altra parte una classe di sventurati la quale in questa lotta per la civiltà non regge alla pugna e, soccombendo, discende in modo rapido, precipitoso, per una decadenza sempre più profonda, in una costituzione inferiore.»

«La società non rinuncia all'avvenire che l'attende per le vittime che si lascia dietro nel suo sviluppo, ma sente di avere in se stessa cause di decadenza che fa duopo eliminare, domanda all'igiene che voglia tutelare la prosperità.»

È questa, a parte un saggio di Cesare Lombroso del 1856, la prima esposizione sistematica di una possibile relazione tra l'incremento della follia e lo sviluppo della civiltà: alla complessità della vita sociale fa riscontro una maggiore complessità della vita mentale ed emozionale interna. Sottesa a questa affermazione si riscontra un'ipotesi di stampo organicistico, consono alla cultura dell'epoca e alla diffusione, all'interno degli ambienti scientifici, delle nozioni della termodinamica: voglio riferirmi ai modelli dell'energetismo e a quelli ancora precedenti sull'eccitabilità dei tessuti, a seguito dei quali il fisiologo John Brown con i suoi *Elementa medicinae* (Edimburgo, 1780) per primo aveva esposto la nozione, storicamente «troppo» fortunata, dell'irritabilità/esaurimento del sistema nervoso.

Dopo queste premesse, Adriani volge il suo interesse principalmente agli interventi profilattici, quelli che possono promuovere, a suo avviso, un più completo e perfetto sviluppo cerebrale.

Questi interventi concernono principalmente la buona alimentazione, l'educazione fisica e gli esercizi del corpo che possono combattere la mollezza delle abitudini, l'educazione dei sensi, la funzione del linguaggio come trasmissione ideale di educazione morale e intellettuale, la riforma dell'istruzione, impigliata in vecchie consuetudini e pregiudizi che sono in opposizione alle deduzioni della scienza sperimentale, l'opportunità di separare l'insegnamento etico da quello religioso il quale «incendia le menti con il misticismo, paralizza la vita, produce superstizioni, terrori, fantasmi», l'indagine sperimentale dei metodi pedagogici applicati e l'intervento del medico nelle scuole.

Tesi simili verranno sostenute da Cesare Lombroso, in scritti di pochi anni seguenti, accusando complessivamente la scuola italiana e i programmi scolastici di «archeofilia».

In breve, compito dello freniatra, secondo Adriani, è quello di contribuire a una educazione che si informi alle regole igieniche e alle leggi della fisiologia, e che si avvalga di indagini statistiche per conoscere l'influenza che la vita civile opera sulla salute mentale.

Negli interventi e nel dibattito si evidenzia una posizione di consenso verso quanto esposto da Adriani. Alcuni come Buonomo sostengono una linea ancora più incisiva: «I medici devono avere l'ispezione delle scuole; studiare una educazione individuale da applicarsi a ciascuno, perché ciascuno porta la sua inclinazione, per eredità o altre cause.»

Altri come il clinico Achille De Giovanni, il fisiologo Luigi Luciani, il freniatra Enrico Gonzales si soffermano sull'opportunità di moltiplicare e differenziare i dispositivi istituzionali e le tecnologie d'intervento deputate alla terapeutica e alla repressione: istituti e collegi speciali per l'educazione degli alunni predisposti alla pazzia, con tendenze alla criminalità e con caratteri eccentrici, pur cercando di non confondere i disgraziati e gli infelici con coloro che hanno tendenze malvage.

L'unico a intervenire sulla questione centrale della relazione civiltà-follia è Andrea Verga, il quale avanza alcune ipotesi alternative a quelle esposte da Adriani: l'aumento della pazzia può essere apparente e «dipendere dalla riforma dei manicomi, per la quale è fatta ora possibilità alle famiglie di sbarazzarsi dei loro pazzi e resa minore la ripugnanza a collocarveli».

«L'aumento dipende in parte anche dai progressi della psichiatria, pei quali ora si riconoscono per pazzi degli individui che un tempo si ritenevano semplicemente stravaganti od eccentrici, degli individui che un tempo finivano i loro giorni nel carcere o sul patibolo, mentre altri godevano della pubblica venerazione e persino dell'onor degli altari.»

Verga, pur ammirato dalla relazione del ben più giovane Adriani, è certamente molto meno propenso ad attribuire una qualche forma di responsabilità, seppure indiretta, alla civiltà: per quest'ultimo la società moderna costituisce un indubbio progresso in quanto disvelerebbe soltanto forme di malattia manifesta (lo scemo del villaggio) o latente (gli eccentrici) le quali precedentemente non cadevano sotto l'attenzione della scienza medica e l'accresciuta sensibilità sociale farebbe sì che vengano attuati gli opportuni interventi.

Tuttavia anche Adriani ha qualche perplessità, ma cautamente indica una strada senza inoltrarvisi: «Ma vi deve essere anche un aumento reale... È importante allora vedere quali sono le forme frenopatiche che più aumentano; questo forse ci può guidare a qualche utile conclusione.»

Il freniatra, grande igienista dell'umanità

Uno spazio particolare merita il pensiero di Cesare Lombroso (1835-1909) per la risonanza che ha avuto, per la organicità totalizzante della sua teoresi, per quanto ha influenzato, in bene o in male, lo sviluppo della psichiatria italiana, della criminologia e del diritto. Le matrici culturali del suo pensiero sono offerte dagli studi giovanili etnografici, dallo storicismo vichiano e antirousseauiano, dalla lettura del biologo Ernst Haeckel e del fisiologo Jacob Moleschott, che assieme alla conoscenza e alla diffusione dei testi di Darwin, lo condurranno a una sorta di «evoluzionismo ecclettico».

Il primo scritto nel quale Cesare Lombroso affronta la tematica civiltà-follia è del 1856: «Influenza della civiltà nella pazzia e della pazzia nella civiltà». Il giovane pensatore (ha solo 21 anni) rileva che la pazzia non è una malattia recente; essa infatti si manifesta già tra i barbari, nel Medioevo, tra le popolazioni extraeuropee.

«La *statistica* mostrerebbe che l'agglomero delle popolazioni aumenta il numero dei pazzi, la *storia* invece oppone che l'agglomero, quando è frutto di vera civiltà, impedisce la pazzia d'imitazione.» La vicinanza dei manicomi, delle autorità tutorie, parenti e vicini argina forme d'imitazione o di contagio.

Nell'interno di un cortile di un manicomio, la rappresentazione della follia è stata colta nell'abbandono dei sensi, nel raptus mistico, nel pianto, nel riso, nel dipinto di Silvio Rotta (Venezia, 1853-1913) intitolato *Manicomio* o *Morocomio*. L'opera del 1895, misura 200 per 300 centimetri ed è conservata alla Galleria nazionale d'arte moderna di Roma. Nella pittura di Silvio Rotta, artista sensibile ai temi melanconici e desolati della condizione umana, si sposano realismo sociale e ideologia cattolica. Dall'Unità d'Italia all'inizio secolo, altri pittori italiani, per sensibilità sociale o per esperienza più o meno diretta, raffigurano il tema del folle e della follia, ora secondo i canoni del pietismo cattolico, ora secondo i moduli del romanticismo tradizionale (Giacomo Grosso), o quelli del verismo sociale (Telemaco Signorini), del simbolismo (Domenico Morelli), del futurismo divisionista (Giacomo Balla) o del ribellismo scapigliato (Lorenzo Viani).

Per quanto concerne le cause fisiche, la civiltà appare impotente. Per quanto concerne le cause morali, ben più incidenti, la civiltà forse accresce di poco i fattori predisponenti.

Lombroso prende in esame tre cause morali principali. In primo luogo l'abuso intellettuale: ma occorre considerare che la divisione degli studi e dei lavori, che il progresso va diffondendo, tende a distribuire l'esercizio a seconda delle forze. In secondo luogo, le passioni: il progresso influisce sul loro sviluppo, sulla loro direzione, ma non le moltiplica. Indebolendo la forza di imitazione delle masse e il prestigio degli alienati, rende impossibile la follia epidemica. Infine, vi sono i bisogni: la civiltà aumenta il numero dei bisogni, ma aumenta anche i mezzi per soddisfarli, con nuove invenzioni e facilità delle comunicazioni.

Nello scagionare la civiltà da ogni possibile causa di incremento della follia, sono qui implicite le tesi sviluppate da Lombroso negli anni successivi sulla follia e sul crimine come fenomeni di regressione atavica, a primordi delle fasi di sviluppo individuale e della specie. «Dove veramente influisce la civiltà è nella forma della pazzia. La pazzia si modella sempre sull'immagine della civiltà in mezzo alla quale imperversa.»

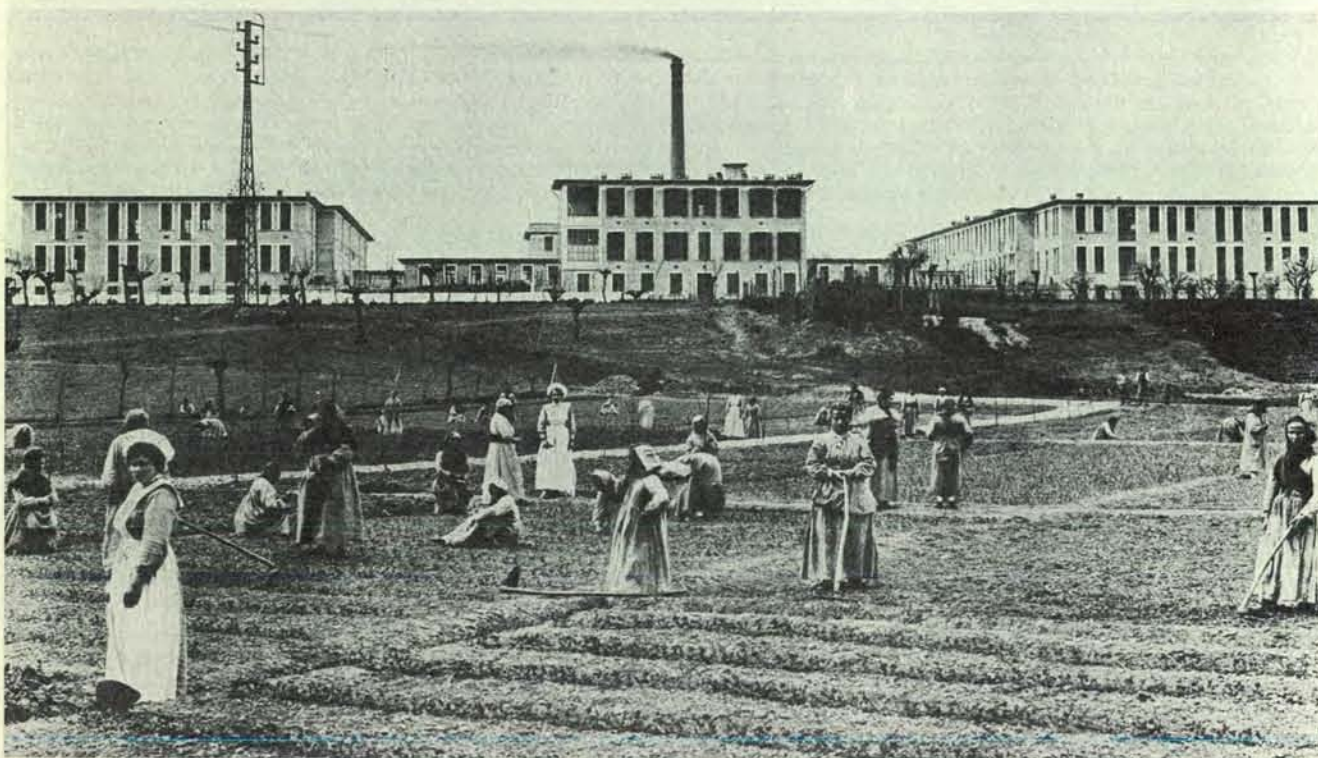
Ma in questo caso Lombroso fa riferimento alle manifestazioni più superficiali della follia (quelle sintomatologiche) e usa il termine civiltà non per indicare quanto concerne lo sviluppo, bensì come sinonimo di cultura: il complesso delle strutture sociali, politiche, economiche, culturali di una data epoca.

Così, in un lavoro del 1860 intitolato «Frammenti medico psicologici, Variazioni della pazzia secondo gli individui, i tempi e le razze», riprende la tesi esposta, documentando come i contenuti delle allucinazioni variano dal Medioevo all'epoca contemporanea, a seconda delle razze, della religione, della cultura.

Praticamente, seppure caratterizzata da toni un po' più pessimistici, questa tesi resta immutata nello sviluppo del pensiero di Lombroso.

Nel 1895 (egli ha 60 anni) riprendendo l'argomento ne «La pazzia nei tempi antichi e moderni», rileva che nell'epoca attuale si presentano nuove forme (paralisi generale progressiva, «epilessia psichica», alcoolismo, morfinismo, cocainismo, isteria e nevrastenia...) alcune delle quali sono certamente caratterizzate da una attività febbrile che fa lavorare il pensiero molto più di quanto la natura non consenta.

Ma a suo avviso la differenza principale consiste nel fatto che nei tempi antichi la follia assumeva carattere epidemico-collettivo; nei tempi moderni e civili, l'avanzare della civiltà con la varietà delle sensazioni e delle idee fa sbocciare il sentimento dell'individualità e permette di rado che intere masse siano predisposte e impressionabili dallo stesso movente.

Un discorso analogo vale anche per altre manifestazioni, quali il crimine: «Il delitto più moderno, che cresce a mano a mano che calano gli omicidi, è quello di frode e bancarotta, e corrisponde all'andarsi sostituendo la frode e l'astuzia alla violenza nella lotta per la vita» (*Delitti vecchi e delitti nuovi*, 1902). Il bancarottiere, ad esempio, manca delle caratteristiche disarmonie fisiche proprie

Il valore terapeutico del lavoro, non un fattore di produzione economica per l'istituzione, bensì una possibilità di recupero del malato, faceva parte integrante del progetto medico sulla follia: il manicomio come un ambiente artificiale, ideato da un gruppo di scienziati competenti. All'interno del «trattamento morale», il lavoro assumeva il ruolo di conservatore della salute, sedativo delle idee irreali, stimolante della volontà, revulsivo della follia, ricostituente del corpo: da esso poteva provenire ordine, disciplina, quiete e contentezza per gli uomini come per l'istituzione. Lavoro per tutti, dunque, secondo le varie disposizioni personali, della malattia e della cultura sociale di provenienza, sotto lo sguardo vigile e benevolo del medico capo che viveva con la sua famiglia all'interno dell'istituzione. Dopo la metà dell'Ottocento, nell'età d'oro dell'alienismo italiano, in molti manicomi nascono le colonie agricole e l'istituzione sembra quasi diventare un sereno villaggio autosufficiente; gli ammalati sembrano scoprire antiche e sopite vocazioni e i freniatri sono orgogliosi dei loro figlioli: pochi sono i veri oziosi, quelli che per mala volontà e cattiva abitudine rifuggono il lavoro. Ma verso la fine secolo, la colonia agricola ha un crollo: l'aumento dei degenti e le nuove tecniche lavorative riducevano la possibilità di occupazione di coloro scarsamente qualificati o difficilmete controllabili. Le fotografie sono «la calzoleria» *(in alto)* e «la colonia agricola - gli orzi» del manicomio di Torino-Collegno.

dei rei-nati, che delinquono in modi più ferini e violenti.

Gli studi di Lombroso sul delitto politico in rapporto alla civiltà permettono di cogliere qualche altro aspetto della complessità del suo pensiero, delle sue matrici storicistiche, della missione professionale o del ruolo sociale del medico freniatra in fine secolo.

Nel 1890, in collaborazione con l'avvocato Rodolfo Laschi pubblica *Il delitto politico e le rivoluzioni*; nel 1895 *Gli anarchici*.

La tesi di Lombroso è brevemente riassumibile: viviamo in una epoca statica, istituzionalizzata, di corruzione del governo e di avidità della classe più abbiente. Nell'incapacità di avviare un processo di riforma sociale che intervenga sui mali profondi della nostra epoca (pellagra, alcoolismo, superstizione, ignoranza scolastica...) nasce un movimento come quello anarchico che vorrebbe promuovere un progresso repentino, ma in realtà è regressivo, verso un'epoca preistorica, «prima che sorgesse il pater familias», dice Lombroso.

Ma i governanti, affetti dallo stesso misoneismo (odio del nuovo) così radicato nel selvaggio, nel bambino e nella donna, hanno tradito il loro mandato: «In un paese - infatti - in cui le riforme politiche vanno di pari passo colle aspirazioni del popolo, le sommosse sono poche o nulle.»

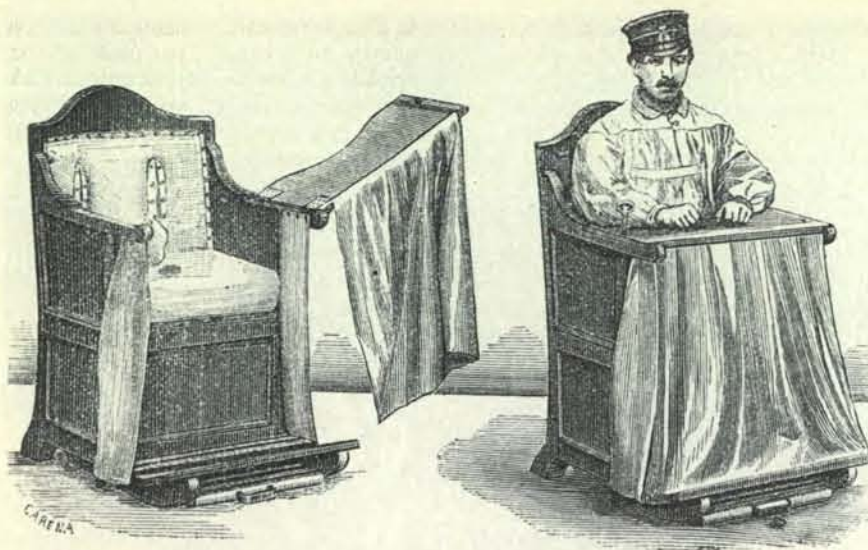
L'interesse per il delitto politico non è una novità di Lombroso, ma ha illustri precedenti in altri freniatri italiani e stranieri. Ma Lombroso radicalizza in qualche modo il discorso: applica le categorie della scienza freniatrica ai rivoluzionari come ai governanti.

All'interno dei primi distingue coloro che presentano sia nei tratti somatici che nel comportamento ferino, un tipo criminale (come Marat, Riel, Hölden, Ravachol...) dai veri rivoluzionari (come Corday, Mirabeau, Cavour, Smisky, Solov'ev, Bakounine, Caserio...) che presenterebbero un tipo perfettamente normale, anzi persino più bello del normale, come tutti gli uomini animati da grandi ideali, distinguendo in questo modo le ribellioni dalle rivoluzioni.

Allo stesso modo, distingue tra i governanti quelli affetti da misoneismo, che regrediscono alla paura del nuovo caratteristica del selvaggio, del bambino e della donna, da quelli mossi da neofilia.

Se è vero che Lombroso era stato negli anni precedenti uno degli illustri e combattivi rappresentanti del movimento igienista, avendogli l'ufficio militare nell'esercito aperto la conoscenza diretta del Mezzogiorno, dei problemi sanitari e igienici che travagliavano le diverse popolazioni e regioni della penisola e avendo successivamente sviluppato studi sulla pellagra e sul cretinismo (indipendentemente dai risultati conseguiti), il filone di indagine al quale approda verso il 1890 è qualche cosa di più e di diverso: il freniatra si erge a «Grande

L'onesta ricreazione dello spirito e del corpo è parte integrante della terapia morale: lo svago come impiego del tempo libero, il lavoro come mezzo di elevazione; ambedue moderatori delle passioni, promotori di ordine e disciplina esterna e interna. Una palestra ove fare ginnastica, un grazioso teatro sul quale i malati si esibiscono al pubblico recitando commedie, una piccola orchestra, la biblioteca dei degenti, un biliardo...qualche uscita dall'istituzione per andare all'opera... Il manicomio si costituiva così sul modello della vita sociale esterna, anzi di una vita «ideale», tranquilla, dove la disciplina garantiva la salute; un mondo nel quale ognuno aveva il suo posto e il suo compito. Solo chi non riusciva ad adattarsi, ancora una volta doveva venire isolato; anche all'interno dell'istituzione si riproduceva il meccanismo dell'espulsione vissuta all'esterno. La fotografia, «udizione radiofonica in teatro», come quelle della pagina a fronte, è tratta da *Il Regio Manicomio di Torino nel suo secondo centenario* (Torino, 1928).

Tra il dispiegarsi di istituzioni e strutture diversificate per gestire i problemi della follia, marginalità e devianza, i patrocini per i dimessi rappresentavano la punta più avanzata della psichiatria filantropica. Promossi a partire dal 1870, tutelavano i pazienti dimessi attraverso un aiuto finanziario, una possibilità lavorativa e una vita comunitaria, affinché questi potessero passare gradualmente dall'istituzione più chiusa del manicomio alla vita sociale e le eventualità di recidive fossero il più possibile contenute. Immagine certa di un lavoro freniatrico nel quale le finalità mediche, sociali e filantropiche trovavano confluenza. La fotografia ritrae Vitige Tirelli, direttore del manicomio di Torino, in visita al convalescenziario di Via Biella, ed è stata riprodotta per gentile concessione della Collezione privata Karen e Bruno Bruni.

Dalla fine del XVIII secolo, Vincenzo Chiarugi in Italia, Philip Pinel e Jean-Étienne Esquirol in Francia, William Tuke e John Connolly in Inghilterra scrivono e lavorano per abolire o ridurre la contenzione nei manicomi. Il trattamento morale prevedeva comunque anche un sistema di premi e di punizioni e il dibattito sulla contenzione è presente per tutto l'Ottocento sia come problema di frenare gli eccessi del male quando potevano condurre ad automutilazioni o a pericolo per altri degenti e personale, sia come problema della violenza del medico a fin di bene (la sonda per far mangiare...). In fine secolo, con l'aumento dei degenti, mezzi di contenzione, seppure umani, vengono utilizzati: nell'inchiesta del 1909 di Giuseppe Antonini, su 50 manicomi, 33 risposero che era in vigore il *no-restraint* relativo. Dopo la guerra mondiale, il dibattito sul *no-restraint* e sull'*open-door*, che aveva anche promosso il nascere delle prime scuole per infermieri, si smorza e di esso non si ha più traccia. Nella fotografia «poltrona per dementi paralitici ed agitati» dell'architetto Carena, in uso al manicomio di Torino dal 1884.

educatore e igienista», per una profilassi e una terapeutica della *ratio* e della vita, colui che vuole indicare al politico come meglio condurre e guidare la vita collettiva, per una società regolata da norme che si richiamino ai principi della fisiologia.

Di qui una serie di scritti per una terapeutica della vita politica: «Le piaghe d'Italia» (1893), «Il momento attuale» (1903). In essi Lombroso indica i pericoli del parlamentarismo, del suffragio universale, espone proposte di decentramento del potere pubblico, l'istituzione di una specie di tribunale del popolo come antidoto alle clientele parlamentari, critica la politica colonialista ed espone radicali posizioni antimilitariste, denuncia la legislazione liberticida del *Codice Zanardelli* (entrato in vigore l'1 gennaio 1890) che «aveva tappato la bocca a tutti gli onesti», si scaglia contro lo spirito burocratico che soffoca ogni riforma, fa proposte per l'organizzazione di una polizia «scientifica», per l'igiene della giornata lavorativa di otto ore, denuncia la deleteria influenza del clero, fuori dal settore morale-religioso, critica l'«archeofilismo» dell'insegnamento scolastico e si pronuncia a favore dell'istruzione laica, obbligatoria e gratuita.

Credo che sarebbe erroneo interpretare gli scritti politici di Lombroso come qualcosa di privato, dell'uomo, del libero pensatore o del socialista. Essi nascono in nome della scienza e della scienza medico-freniatrica, e in essi egli cerca di applicare proprio le categorie della sua disciplina ai fenomeni sociali: «Il sentire i proprii tempi, il partecipare agli eventi che accalorano le masse...è una necessità che si impone anche allo scienziato» (come si legge nella presentazione a «Il momento attuale»).

Che questo atteggiamento, che ho provato a indicare come quello del «grande igienista», si connoti come indagine clinica sperimentale negli studi di un fisiologo come Angelo Mosso, come inchiesta sulla mentalità delle classi povere, o banalizzante e terroristica indagine (eziologia, diagnosi e prognosi) della mentalità di *fin de siècle* negli scritti del famoso medico giornalista dell'epoca Max Nordau, non ha importanza. Quanto mi pare opportuno rilevare è l'atteggiamento mentale con il quale i freniatri di fine secolo si prefiggono di partecipare alla vita nazionale italiana.

Il nuovo secolo: la «scienza normale»

Con l'inizio del secolo ogni problema sembra trovare la più opportuna definizione: si entra nella fase descritta da Thomas Kuhn della «scienza normale».

Il Governo sembra ora accettare in pieno quanto per anni reclamato dalla Società freniatrica. Al congresso del 1901 tenutosi ad Ancona, le parole del prefetto della provincia, Commendator Ovidi, a nome del Ministro dell'interno, Giovanni Giolitti, riproducono esattamente quanto esposto da Verga circa venti anni prima: «Il cammino che la vostra scienza ha percorso in questi ultimi anni risplende di tanta luce che cieco è chi non lo vede. Quanta differenza, infatti, dai giorni della mia fanciullezza a oggi nei modi di intendere e di trattare quel gran flagello dell'umanità, quasi negazione di Dio, che è la demenza!

«Allora, quando si toccava di questo argomento, era un continuo parlare di frusta, di nerbo, di aguzzini: le cose più tetre, scherni e percosse, si raccontavano della vita interna dei manicomi... Ora è tutt'altra cosa: e ciò principalmente per effetto della vostra scienza serena e pietosa; ed ogni nuovo manicomio che sorge fa di ciò ampia testimonianza. Ché ogni nuovo manicomio è costruito con intelletto d'amore...

«La vostra scienza...coadiuvata dalle scienze e dalle arti affini penetra, per di più, molto addentro nelle manifestazioni tutte della psiche... Interessa lo statista, ma più ancora l'avvenire stesso della società... Così, tra i tanti servigi che potete rendere alla nostra Italia vi sia, insieme a quello della educazione e dell'istruzione, vivamente raccomandato il problema della nostra delinquenza...».

A queste parole fa eco la relazione di Augusto Tamburini, allora direttore del manicomio di Reggio Emilia: «Le conquiste della psichiatria nel XIX secolo ed il suo avvenire nel secolo XX», in cui si legge: «Il secolo XIX può ben vantare, fra le immense e gloriose sue conquiste, anche quella d'aver costituito a scienza la psichiatria e d'aver compiuto i più grandi progressi nel trattamento umanitario dei pazzi.

«La psichiatria scrutava e svelava la mente e l'azione di molti degli uomini che hanno esercitato la maggiore influenza negli avvenimenti mondiali...poneva sotto una nuova luce gran parte dei fatti della Storia universale, spiegandone la genesi, il decorso, gli effetti...ha fornito di nuovi e copiosi elementi alla letteratura e alle arti...

«Ha dimostrato su quali basi informi ed erronee fossero fondati sinora i sistemi ed i metodi della pedagogia...ha posto le salde radici...di una Patologia e di una Terapia sociale.»

Ma la cura esatta del meccanismo intimo, del processo patogenetico della pazzia, una cura razionale ed efficace, porterà sicuramente alla diminuzione della frequenza e al conseguente sfollamento di manicomi - afferma il relatore.

Del connubio ormai avvenuto tra psichiatria e potere politico sono testimonianza le 17 pagine della relazione introduttiva di Leonardo Bianchi, medico e legislatore, al disegno di legge del 1904.

«...la pazzia aumenta paurosamente, e il legislatore...non si può disinteressare ... Il complicato meccanismo della civiltà odierna, i moltiplicati e diversi rapporti sociali di ciascun individuo nell'ambiente sociale, la somma di lavoro che riprende lenta nel miraggio del godimento, il cresciuto numero di desideri e delle disillusioni, il valore della vita aumentato e la intensificazione della lotta per essa richiedono un potenziale intellettivo di gran lunga superiore a quello che fosse necessario non più che mezzo secolo e anche trent'anni fa... Il pazzo, tutto sommato, non è che un debole nella lotta per l'esistenza... Il folle che ingombra la società simboleggia, nell'organismo sociale, quello che rappresentano le tossine, le infezioni nell'organismo individuale... Poiché la scienza è riuscita a classificare i delinquenti e a riconoscere la particolare fisionomia psicologica di ciascun gruppo, è bene che il legislatore utilizzi questa nuova conoscenza in un regolamento per i manicomi.»

L'analisi non può apparire che sommaria, ma forse non interessa una analisi del fenomeno. La genericità delle norme della legge può permettere di capire molto di più: il fenomeno è ormai incanalato e vengono concessi ampi spazi di discrezionalità al gruppo professionale degli psichiatri, preposto alla cura e all'assistenza. Un illimitato potere che qualche giurista, come Luigi Lucchini, rileva con durezza e perplessità nella discussione del progetto di legge.

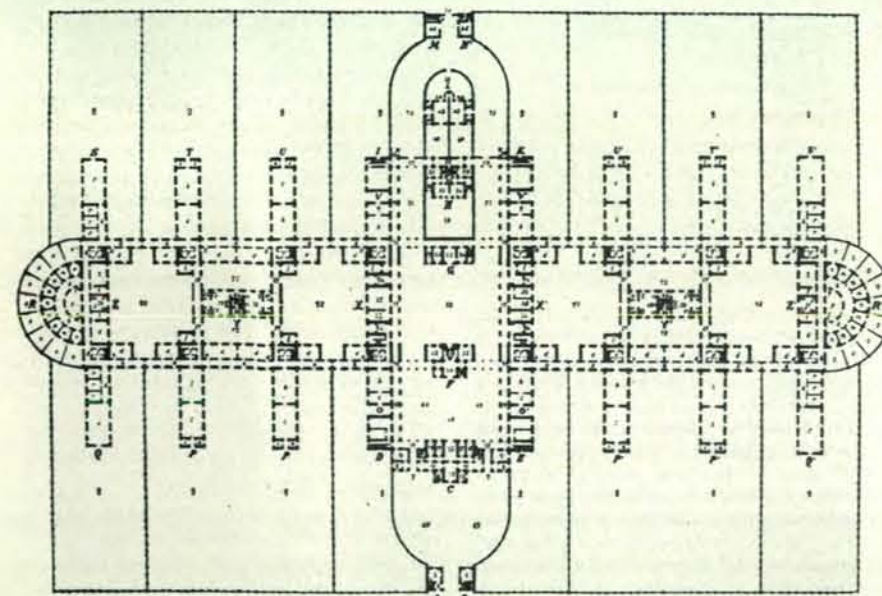
Il codice teorico

All'undicesimo congresso della Società freniatrica italiana, il medico Sante De Sanctis, uno dei padri fondatori della psicologia italiana, cerca di fare il punto sulla esuberanza di classificazione che dal primo congresso degli scienziati italiani a Pisa nel 1839 sono state successivamente proposte. Egli rileva come vari criteri tassonomici (anatomo-patologico, biologico, psicologico, etio-patogenetico, sintomatologico, prognostico) si siano via via intrecciati nelle classificazioni sinottiche proposte, ma osserva come sia prevalso in genere un carattere descrittivo in un assoluto difetto di sistematicità. Rileva come la tendenza di numerosi giovani psichiatri converga attualmente verso la classificazione del neuropsichiatra tedesco Emil Kräepelin.

La proposta di classificazione di Lombroso non viene neppure più menzionata, sebbene abbiano lavorato attorno a essa per decenni numerosi freniatri italiani: è, come si dice in termini moderni, una classificazione storicamente perdente. Essa era sostanzialmente costituita su quattro criteri (nato, per passione, per abitudine, occasionale) che potevano essere applicati ai più svariati fenomeni (follia, crimine, prostituzione, delitto politico, omosessualità...).

La classificazione storicamente vincente sarà in effetti quella del trattato di Kräepelin (l'edizione tedesca del 1883 viene pubblicata in italiano dalla Vallardi nel 1897, ma già nel 1899 è necessaria una seconda edizione) che è basata su un criterio prognostico per distinguere le varie entità morbose; raggruppa per la prima volta ebefrenia, catatonia e demenza paranoide nell'unica classe della «demenza precoce», mania e melanconia nella classe della «psicosi maniaco-depressiva».

La classificazione di Kräepelin estremamente organica nel distinguere sintomi, sindromi, entità morbose, e il territorio della psicosi da quello della nevrosi viene presto adottata, promuovendo così un primo processo di omogeneizzazione del linguaggio psichiatrico.

È facilmente comprensibile come per una comunità di scienziati sia un momento storico particolarmente importante quello in cui si trova un codice

Le strutture edilizie manicomiali nascevano, nel caso di ospizi nuovi, a immagine speculare dell'organizzazione del sapere psichiatrico: un tecnico, il freniatra, chiedeva a un altro tecnico, l'architetto, le soluzioni più funzionali, la segmentazione dello spazio in una serie di unità staccate, corrispondenti ai blocchi di terapia. Il manicomio di Voghera (1876) per progetto degli architetti Monti e Savoldi su indicazioni di Augusto Tamburini proponeva il modello asiliare-ospedaliero: un sistema a bracci legato, circondato da giardini. Il manicomio di San Lazzaro, di Reggio Emilia, la cui fama era tale che vi ricorrevano anche le famiglie più agiate, consisteva di 24 fabbricati, indipendenti gli uni dagli altri. Costruito in epoche diverse e per opera di ristrutturazioni successive, riproponeva nelle sue parti l'impronta cronologica di ogni epoca. Un villaggio praticamente, con un padiglione centrale per l'amministrazione, la direzione medica, un salone per concerti e feste, diversi villini disseminati, le officine, la colonia agricola... Qui a fianco è riportata la planimetria del manicomio di Voghera; in alto, una veduta del manicomio di Reggio Emilia.

ANNO	POPOLAZIONE	NUMERO DEI PAZZI INTERNATI	INCIDENZA PER 1000 ABITANTI	FONTE
1874	23 967 736	12 210	0,5	VERGA
1877	28 010 695	15 173	0,5	
1880	28 524 399	17 471	0,6	
1883	29 010 652	19 656	0,6	DIREZIONE GENERALE DI SANITÀ
1888	30 497 610	22 424	0,6	
1891	30 686 334	24 118	0,7	LOMBROSO E TAMBURINI
1896	31 195 697	29 631	0,9	TAMBURINI
1899	31 479 217	36 931	1,1	DIREZIONE GENERALE DELLA STATISTICA
1908	33 506 900	45 009	1,3	
1910	34 565 804	51 215	1,4	TAMBURINI E ANTONINI
1914	35 238 997	54 311	1,5	VIDONI

Quando lo Stato riformista del Settecento cerca di sostituirsi nell'assistenza sociale ai tradizionali sistemi di beneficienza e carità espletati dagli ordini religiosi, per la prima volta si cerca di censire i pazzi. All'inizio dell'Ottocento la statistica, da tecnica dell'amministrazione, viene applicata alle scienze dell'uomo e diviene tecnica morale, espressione del bisogno d'ordine e di efficienza; più tardi, con l'unità dello Stato nazionale, strumento per inventariare e conoscere le condizioni e i beni d'Italia. In fine secolo il dispiegarsi di strutture deputate alla follia e alla emarginazione appare ingente e allo stesso tempo sempre insufficiente a fronte di un aumento costante del numero delle persone istituzionalizzate. Vi sono 55 manicomi pubblici, tre giudiziari, 30 privati, 51 altri istituti, in buona parte dislocati nell'Italia settentrionale. La popolazione italiana è aumentata di un terzo, mentre la popolazione degli alienati è aumentata del triplo, nonostante che abbiano un peso meno rilevante le grandi sindromi di matrice carenziale, come la pellagra e il cretinismo. La cartina riporta i manicomi del Regno d'Italia, mentre la tabella riproduce i dati del censimento dei pazzi ricoverati in Italia dal 1874 al 1914.

teorico all'interno del quale convergere, e il trattato di Eugenio Tanzi del 1905, il primo sistematico della psichiatria italiana, praticamente si uniforma nella sua ossatura teorica alla nosografia tedesca.

L'emergere dello «sguardo» nel campo dei segni e dei sintomi consente di strutturare un campo concettuale ove l'individualità completamente osservata si perde in favore della molteplicità categoriale percorsa dai fatti individuali concreti. Spiega Michel Foucault in *Nascita della clinica* come il sintomo soggettivo divenga un segno sotto uno sguardo sensibile alla differenza, alla simultaneità, alla successione e alla frequenza.

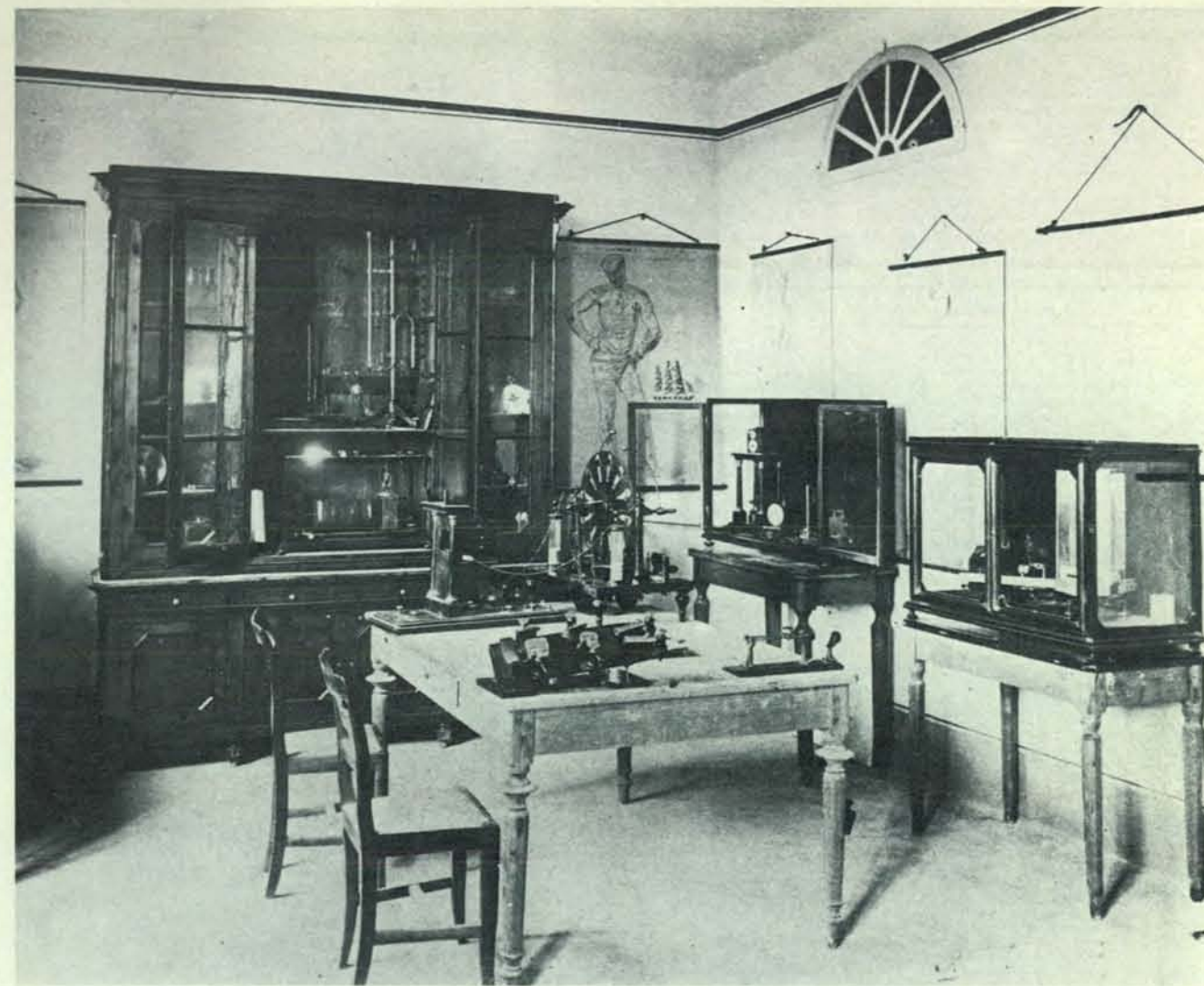
Le tecnologie d'intervento

Tutto il dibattito di fine Ottocento era incentrato sulla opportunità di ampliare e diversificare le strutture deputate al trattamento della follia e della marginalità, per un più ampio ventaglio di interventi profilattici, terapeutici e repressivi, per un ampliamento di quelle che Foucault ha definito «eterotipie», «luoghi anomali ed introflessi che ininterrottamente agiscono sui soggetti custoditi, ma che interrompono la continuità dei legami... Mai peculiari ad alcun contesto, ma ovunque localizzabili nelle aree che la città lascia ai propri confini».

Così la vita interna di queste strutture tendeva a plasmarsi sul modello della vita sociale esterna, anzi di una vita «ideale», tranquilla, nella quale ognuno doveva avere il suo posto e il suo compito da svolgere: la «cura morale» di inizio secolo, sorta sul tentativo di educare al controllo delle passioni, veniva a modellarsi sulla funzione simbolica del lavoro, con la funzione terapeutica di allontanare il pensiero dal delirio o dalle idee dolorose, e di disciplinare il disordine interno.

Il manicomio sembra così trasformarsi in un idilliaco villaggio autosufficiente ove ognuno è deputato alle proprie tradizionali mansioni. Le attività lavorative, assieme alla presenza all'interno dell'ospedale della figura carismatica del «direttore» che vi abitava con la sua famiglia, assumevano il compito di plasmare «dolcemente» gli alienati all'universo asiliare. Se questo non avveniva, si riproducevano meccanismi di esclusione e di repressione che erano caratteristici della vita esterna.

All'inizio del secolo, allorquando gli psichiatri hanno raggiunto, almeno in buona parte quanto si prefiggevano (da una indagine del 1898 risultano 43 manicomi, 32 case di cura private, tre manicomi giudiziari, 16 cliniche psichiatriche universitarie, senza contare una miriade di istituti per minori), il tema del confronto di idee si sposta sulle modalità di gestione delle istituzioni.

Tale è il significato di buona parte delle relazioni del congresso del 1901 e degli anni seguenti sulla tecnica manicomiale, l'opportunità dell'*open-door* e del

Il laboratorio di psicologia di fine Ottocento era caratterizzato dagli strumenti per le indagini psicometriche, le ricerche sulla velocità della corrente nervosa e la misurazione del periodo di eccitazione. Nel 1829 Francesco Saverio Hildebrand dell'Università di Pavia aveva presentato un progetto per l'istituzione di una clinica psichiatrica, ma solo nel 1850 Stefano Bonacossa iniziava il primo corso di freniatria presso l'ospedale di Torino: scuole e cliniche psichiatriche trovano luogo nei manicomi, dove cioè si fa clinica e ricerca. In molti ospedali nascono il gabinetto patologico, il museo antropologico-craniologico, il laboratorio di psicologia, di fotografia ecc. È il definitivo tramonto del periodo assistenziale, mentre lo sguardo ordinatore e onnicomprensivo del freniatra tenta di collegarsi alla medicina generale e di valersi delle scienze biologiche fondamentali. Nella fotografia, il laboratorio di psicologia del manicomio di San Lazzaro di Reggio Emilia.

no-restraint (un dibattito introdotto in Italia a opera di Ruggero Tambroni, promotore dal 1890 al 1930 della Scuola psichiatrica ferrarese), esperienze da cui hanno anche origine le prime lezioni agli infermieri, prima manifestazione dell'esigenza di un personale adeguatamente preparato; i provvedimenti da prendersi per diminuire l'affollamento dei manicomi; le prime polemiche sulla gestione dei manicomi giudiziari: gli stessi Lombroso e Tamburini, promotori nel decennio 1870-1880 di questi istituti, a distanza di pochi anni ne diventano critici spietati, sconcertati dalla concretizzazione storica del progetto.

Nello stesso periodo comincia anche a manifestarsi una linea di intervento più «segregante», che tende a ridurre le visite dei familiari, e «spersonalizzante», che vuole privare i pazienti di tutti gli effetti personali. L'oggettivazione della follia diventa più evidente allorquando scompaiono i diari giornalieri del decorso e compaiono le prime cartelle cliniche nosografiche stampate di Lucio Ronconi (nel 1877 medico capo del manicomio di Alessandria), predisposte a una compilazione che prevede la storia della malattia e le misure antropometriche, ma che esclude le notizie familiari, soggettive, sociali.

Accade così che nella società moderna e in un dato periodo storico, dal dibattito che vuole promuovere un dispiegarsi di istituzioni e strutture diversificate per gestire problemi della follia, marginalità e devianza, si sia passati al dispiegarsi di modalità di gestione diverse delle strutture medesime.

Il grande internamento

Nonostante la preoccupazione e le proteste di tutti, alla fine dell'Ottocento il numero dei degenti nei manicomi continua a salire vertiginosamente: da 47 ogni 100 000 abitanti del 1875, si arriva a circa 150 ogni 100 000 nel 1910.

Certamente appare ingenuo imputare l'aumento dei degenti a una diminuzione della diffidenza della gente comune nei confronti del manicomio, che permetterebbe di fare ricoverare i propri cari, tesi sostenuta da molti psichiatri quali Tamburini, Lombroso, Tanzi, Guido Ruata, Luigi Anfosso.

Romano Canosa, studioso contemporaneo, rileva che la curva dell'internamento carcerario segna una netta inversione di tendenza alla fine del decennio

REGIO MANICOMIO DI TORINO

TABELLA NOSOGRAFICA

REPERTO NECROSCOPICO

ETIOLOGIA

ANAMNESI

ESAME PSICHICO

ESAME CLINICO

DECORSO DELLA MALATTIA

La cartella clinica cambia progressivamente: dalla pagina bianca, senza schemi preformati di registrazione dei dati dell'inizio Ottocento e sulla quale veniva annotato un po' di tutto, in indeterminatezza di linguaggio e di concetti teorici, verso metà secolo compaiono le prime cartelle preordinate a cogliere alcuni aspetti del comportamento e a escluderne altri; l'attenzione dei freniatri ai fattori sociali patologici e biologici è tale che in alcuni manicomi viene istituito un diario giornaliero per ogni degente. Poi, gradatamente i modelli teorici e le categorie osservative vanno cambiando: la storia sociale premorbosa e le problematiche del soggetto e della comunità di appartenenza vengono sempre più trascurate, mentre vengono privilegiati i dati somatici e psichici e la cartella clinica viene preordinata per riportare i dati dell'antropometria lombrosiana. La storia della cartella clinica, come strumento nel quale riportare i dati osservativi, viaggia pertanto di pari passo con i dibattiti e i progetti per la classificazione delle malattie mentali e con i primi tentativi per distinguere i sintomi dalle sindromi e dalle eziologie. Nell'illustrazione è riprodotta una cartella clinica completa del manicomio di Torino-città compilata nel 1902.

1881-1890, mentre quella dell'internamento manicomiale continua a salire. Questa ascesa si verifica nonostante che verso la fine del secolo vadano perdendo peso le gravi sindromi di matrice carenziale, come la pellagra.

Inoltre, dopo la legge del 1904, le dimissioni diminuiscono vertiginosamente. Il meccanismo stesso della legge rendeva difficile per lo psichiatra assumersi la responsabilità della dimissione, responsabilità che sentiva di assumersi durante l'Ottocento.

La «civiltà»

Con l'inizio del secolo prevale la tesi che la civiltà è indubbiamente assai più rivelatrice che produttrice di un gran numero di alienati.

La civiltà viene ora intesa come progresso lineare, inarrestabile, intrinsecamente giusto; le sue vittime, la massa crescente degli «alienati», rappresentano così un prezzo pressoché inevitabile, perché «con l'avanzare della Civiltà...è più facile il rivelarsi dei deboli e degli inetti e più sentita la tendenza selettiva della loro separazione» e inoltre perché vi sono «dovunque condizioni genetiche in aperta opposizione alla Civiltà che mantengono o aumentano le aberrazioni psichiche».

La tesi che la malattia mentale è un fatto biologico diventa quella ufficiale della Società freniatrica: il progresso scientifico e l'apertura dei manicomi può finalmente curare quanto precedentemente veniva abbandonato al pubblico ludibrio. Come in questi anni il movimento igienista tende a scientificizzarsi, allo stesso modo l'ideologia psichiatrica abbandona progressivamente le caratteristiche umanitarie degli anni precedenti per sostituirvi un atteggiamento avalutativo e scientificamente neutro.

L'ideologia del progresso ha ormai definitivamente sconfitto gli ultimi retaggi della medicina romantica, che aveva individuato nella civiltà una qualche forma di disturbo; la fede nel progresso diviene fiducia certa nel perfezionamento del genere umano, visione idilliaca e apologetica nei confronti dell'incivilimento. E qualcuno sostiene che colpevolizzare lo sviluppo civile sarebbe un po' come accusare una madre di fare impazzire i propri figli, dopo averli messi al mondo.

Tra le ultime voci che si pronunciano sul dibattito civiltà-follia, quella di Tanzi, nel suo *Trattato* del 1905, appare la più ferma e conclusiva, prima che la tematica scompaia all'attenzione degli psichiatri: «Il cervello è un osservatorio, un archivio, un tribunale, un governo.» Non possono essere considerate cause dell'incremento della follia né lo studio più intenso e generalizzato, né ideali più vasti e audaci, né il dileguarsi della fede religiosa. Forse le crisi economiche e l'inacerbirsi della concorrenza incrementano la follia, ma questi non sono fattori della civiltà, bensì retaggi della barbarie. «La Civiltà vera che non fu ancora attuata è quella esente da crisi e da odiose sperequazioni che esagerano in luogo d'attenuare le differenze naturali tra individuo ed individuo.»

Un progetto politico che era durante l'Ottocento commensale al progetto medico-scientifico sembra scomparire di scena con l'inizio del Novecento, o per lo meno manifestarsi in modo meno aperto e palese allorquando il corpo professionale degli psichiatri ha raggiunto una serie di obiettivi professionali.

Nel contempo lo sguardo osservativo del «grande igienista» sembra spostarsi dal crimine e dal delitto all'emotività. Grazie alla patologia dell'emotività la zona di influenza della psichiatria si amplia ulteriormente: gli abnormi dell'emotività individuale e sociale sono infatti intiere legioni.

La ricerca del dato obiettivo con il quale comprovare e fondare al contempo un modello interpretativo della malattia mentale aveva già avuto inizio con gli studi di fisiognomica e di ritrattistica. Ma la macchina fotografica offriva ora una grande occasione per fissare a permanenza le fisionomie dei malati nelle varie fasi della malattia, costruire un elemento di studio per la semiologia delle malattie mentali e documentare, assieme alle misurazioni craniologiche, come chi viola la norma sociale abbia già violato la norma biologica. Nel 1862 il medico Duchenne de Boulogne, che per primo aveva fotografato i suoi malati, pubblicava *Le mécanisme de la physionomie humaine* e dal 1870 Augusto Tamburini a Reggio Emilia, Prosdocimo Salerio a Venezia, Gaspare Virgilio ad Aversa, Vitige Tirelli, Antonio Marro e Cesare Lombroso a Torino aprivano nei manicomi gabinetti di fotografia e promuovevano una tradizione. Poco dopo sulla base della stessa tecnica, Louis Adolphe Bertillon a Parigi e Salvatore Ottolenghi a Roma introducevano l'uso della fotografia segnaletica giudiziaria. La fotografia consentiva finalmente e su vasta scala di fissare l'identità. La fotografia, una «isterica-passionale», fa parte della Collezione Marro del Museo di antropologia ed etnografia di Torino.

Il prossimo grande terremoto della California

Anche se, per una previsione esatta, è necessaria una migliore conoscenza dei meccanismi che scatenano i terremoti, si stima probabile nei prossimi 30 anni un sisma di grande intensità lungo la faglia di San Andreas

di Robert L. Wesson e Robert E. Wallace

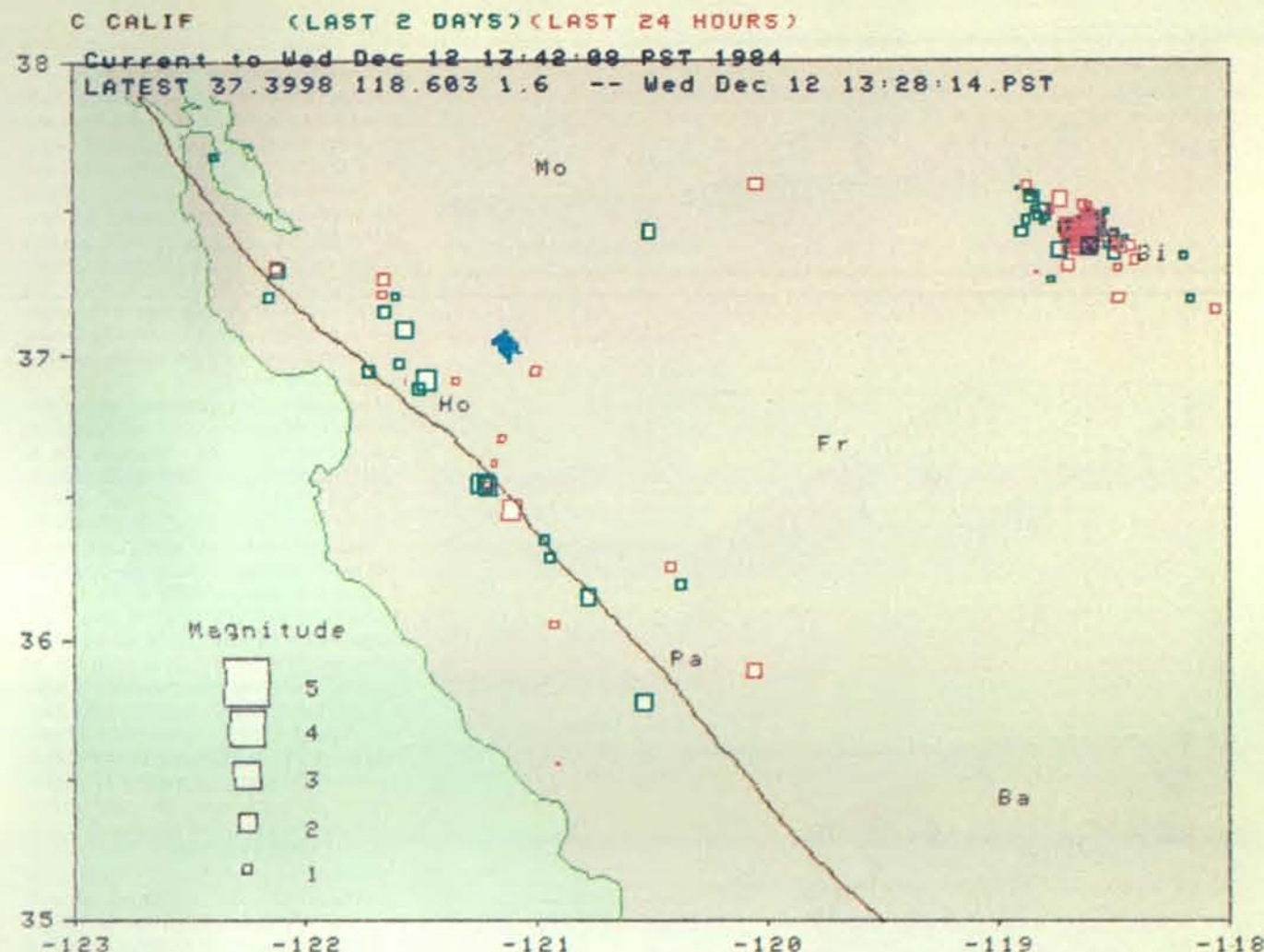
Con la previsione, risultata successivamente vera, dell'eruzione del Mount Saint Helens nel maggio 1980, le scienze della Terra sono entrate negli Stati Uniti nell'era della geologia in tempo reale: lo studio dei processi geologici mentre accadono, a volte minuto per minuto. La previsione in sé fu tutt'altro che perfetta; mancò, in realtà, un allarme a breve termine. Ciononostante, la consulenza dei geologi indusse le autorità governative a livello federale, di stato e locale a istituire una serie di misure precauzionali, fra le quali, per esempio, la chiusura di aree forestali, che senza dubbio ebbero il merito di ridurre la perdita di vite umane e di beni a causa dell'eruzione.

In California l'evento geologico che maggiormente necessita di una previsione esatta - quello la cui distruttività potrebbe essere meglio scongiurata - è un terremoto catastrofico. In termini geologici, la California è un regno di terremoti: essa fa parte dell'Anello di fuoco, la fascia di attività sismica e vulcanica che accerchia il Pacifico. La faglia di San Andreas, che taglia la California da nord a sud, è una delle manifestazioni più evidenti dell'attività sismica della Terra. La California è anche lo stato più popoloso degli Stati Uniti e sede di molte industrie orientate su tecnologie molto avanzate e cruciali per la nazione. In California si trova il 10 per cento circa delle risorse umane e industriali della nazione e l'85 per cento di queste risorse (ovvero l'8,5 per cento del totale nazionale) è concentrato in una fascia di 21 contee lungo il margine continentale all'interno della zona sismica della faglia di San Andreas. Una singola contea - la Santa Clara County, nella California settentrionale - produce un quarto dei semiconduttori della nazione. Essa subì gravi danni per il terremoto che nel 1906 devastò San Francisco.

Nella scia dell'esperienza acquisita in occasione dell'eruzione del Mount Saint Helens, il National Security Council ha assegnato a un gruppo, appositamente costituito, di ricercatori e di enti, fra i quali l'US Geological Survey, il compito di prevedere le conseguenze di un ipotetico terremoto in California. Il gruppo ha ipotizzato vari terremoti potenzialmente dannosi, fra i quali uno simile al grande sisma del 1857, occorso lungo la parte meridionale della faglia di San Andreas. Oggi si valuta che il tempo di ricorrenza di una scossa sismica del genere sia di circa 140 anni e quindi la probabilità che essa abbia luogo in un qualche momento nel corso dei prossimi 30 anni è stimata attorno al 50 per cento.

Secondo l'analisi eseguita, l'ipotetico terremoto potrebbe causare un numero di vittime compreso tra 3000 e 13 000. Il valore più basso corrisponde all'eventualità che il sisma accada alle 2:30 del mattino. La ragione emerge dalle osservazioni su scala mondiale del modo in cui i terremoti agiscono sul tipo di edifici diffusi in California. In sintesi, la popolazione della California è più sicura quando si trova a letto a casa propria. Le case unifamiliari in legno presentano una forte resistenza ai danni strutturali provocati dai terremoti. Le perdite più elevate si avrebbero se il terremoto avesse luogo alle 16:30 di un giorno lavorativo, quando gran parte della popolazione si trova nei luoghi di lavoro o in strada. Anche in questo caso la spiegazione va ricercata nella struttura degli edifici. Le vecchie costruzioni in mattoni non rinforzate da acciaio crollerebbero probabilmente anche nel caso di un terremoto di modesta intensità. Nel terremoto che nel maggio 1983 colpì Coalinga, in California, 30 dei 40 edifici in mattoni della città costruiti prima del 1930 crollarono o subirono gravi danni. Edifici del genere sono comuni nel centro di quasi tutte le città e cittadine più vecchie della California; nella sola Los Angeles se ne contano circa 8000. Essi vengono gradualmente sostituiti, specie a Los Angeles, ma continueranno a esistere in numero significativo ancora per molti anni.

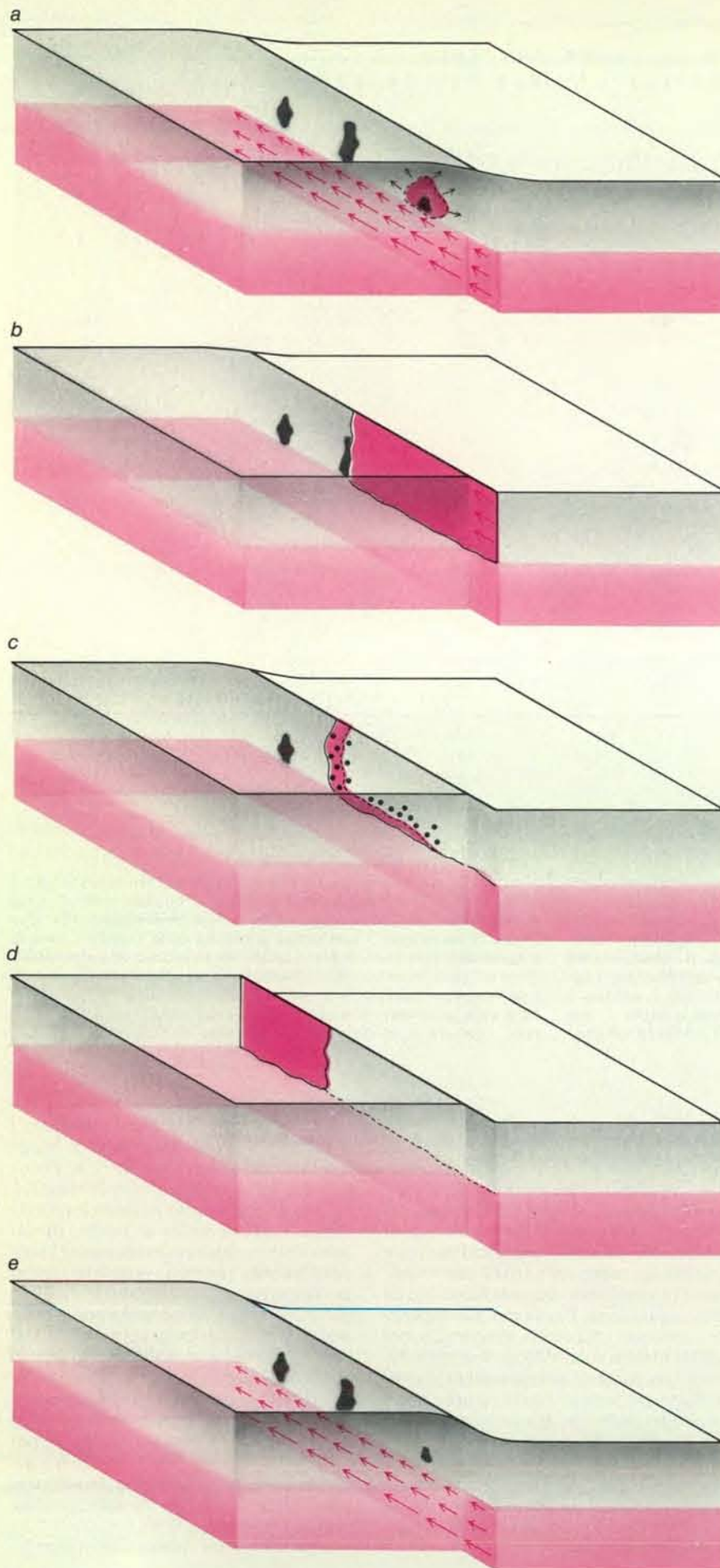
La previsione dei terremoti è al centro della ricerca scientifica di base. Impegnativi programmi di ricerca sono in corso in Cina, in Giappone e in Unione Sovietica. Negli Stati Uniti si spendono ogni anno circa 17 milioni di dollari nel campo della previsione e altri 43 milioni di dollari in un programma di ricerca comprendente studi sulla valutazione del rischio sismico, sulla costruzione di edifici antisismici e sulla natura fondamentale dei terremoti. I progressi nella previsione a breve termine sono più lenti di quanto i ricercatori sperassero una decina di anni fa, quando furono riferiti i primi successi da parte di ricercatori sovietici e americani. D'altra parte vi sono stati progressi confortanti nella previsione a lungo termine e nella comprensione dei meccanismi che scatenano i terremoti. Ora più che mai l'obiettivo della ricerca è quello di imparare a interpretare i segni che indicano l'accumularsi di tensione all'interno della Terra nel quadro globale delle conoscenze di base del processo sismico.

Un primo passo verso la conoscenza attuale dei meccanismi che danno origine ai terremoti è stato compiuto all'inizio di questo secolo, quando si riconobbe nella faglia di San Andreas la causa di forti eventi sismici; era stata la faglia a provocare il terremoto di San Francisco del 1906. Dopo una tale catastrofe Henry F. Reid, della Johns Hopkins University, avanzò l'ipotesi secondo la quale i terremoti sono generati dall'improvviso scivolamento e riassestamento elastico dei blocchi crostali ai bordi di una faglia. In sostanza, egli suggerì che la crosta poteva piegarsi sotto sforzo, poi spezzarsi e raddrizzarsi, liberando l'energia accumulatasi nella deformazione.

La cartina prodotta al calcolatore mostra l'attività sismica nella California centrale durante le 48 ore precedenti le 13:42:08, Pacific Standard Time (PST), di mercoledì 12 dicembre 1984. L'attività è stata rilevata da sismografi, distribuiti in tutto lo stato, collegati a un calcolatore dell'US Geological Survey di Menlo Park. Il calcolatore ha determinato entità e ubicazione dei terremoti; la registrazione è aggiornata ogni qualche secondo. San Francisco, in alto a sinistra, è situata, con Hollister (*Ho*) e Parkfield (*Pa*), lungo la faglia di San Andreas (*linea scura*). Modesto (*Mo*), Fresno (*Fr*) e Bakersfield (*Ba*) sono invece più all'interno. La concentrazione di terremoti in alto a destra fa parte di un'esplosione di attività che ebbe inizio con un sisma di magnitudo 5,7 il 23 novembre 1984 nei pressi di Bishop (*Bi*). Una scossa di magnitudo 3 può essere percepita nelle vicinanze; una di magnitudo 5 può causare danni locali; un terremoto di magnitudo 7 viene avvertito in un'area di centinaia di migliaia di chilometri quadrati e può causare danni estesi; un sisma di magnitudo 8 viene considerato un grande terremoto. L'aumento di un'unità nella scala delle magnitudo equivale a un aumento di 30 volte nell'energia del sisma.

Il modello di Reid, associato ora alla teoria della tettonica a zolle formulata negli anni sessanta, è la base per valutare quando e dove si possono avere grandi terremoti. Nella teoria della tettonica a zolle la maggior parte dei grandi terremoti vengono spiegati come improvvisi movimenti o fratture lungo le faglie che costituiscono i margini delle grandi zolle crostali della Terra. Secondo la teoria, quella di San Andreas fa parte dell'insieme di faglie che segnano il margine fra la zolla del Pacifico e la zolla del Nord America. Le zolle scivolano l'una rispetto all'altra; in particolare, il fondo del Pacifico si muove verso nord-ovest rispetto all'America Settentrionale. Sul lungo termine, la faglia di San Andreas deve adattarsi a uno scorrimento di circa 3,5 centimetri all'anno. La conclusione inevitabile è che improvvisi scorrimenti futuri non potranno che dare origine a terremoti potenzialmente catastrofici.

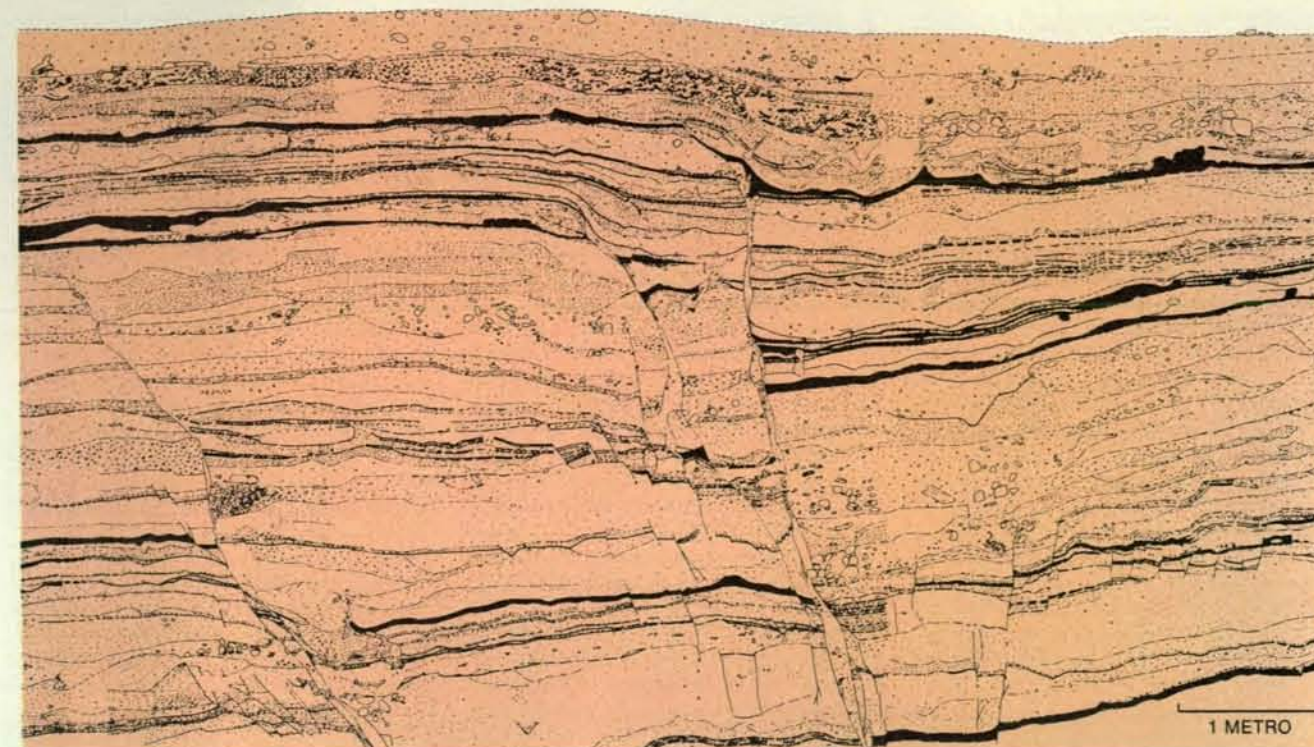
In che modo, precisamente, i terremoti si inquadrano nell'ambito della teoria della tettonica a zolle? Fondamentalmente lo scorrimento fra due zolle procede senza produrre fenomeni sismici a profondità superiori a 10-12 chilometri, dove le rocce sono abbastanza calde da essere plastiche. Più vicino alla superficie, le rocce della crosta sono fragili, cosicché lo scorrimento lungo una zona debole - in particolare una faglia - è controllato dall'attrito. Un terremoto viene prodotto dallo scorrimento improvviso fra blocchi nella parte superiore della crosta. Lo scorrimento può essere avviato da una frattura che si produce in una parte particolarmente resistente (ossia ad alto attrito) della faglia, nota ai geologi come asperità.

Come si possono dunque prevedere i terremoti? La previsione a lungo termine ebbe inizio quando S.A. Fedotov, dell'Istituto di geofisica di Mosca, e K. Mogi, dell'Istituto di ricerca sui terremoti dell'Università di Tokyo, riuscirono a rappresentare graficamente le zone di frattura per tutti i grandi terremoti avvenuti lungo il margine occidentale del Pacifico. Le zone poterono essere dedotte dalla distribuzione spaziale delle scosse successive (o repliche) di ogni terremoto. Fedotov e Mogi notarono entrambi che le zone tendevano a confinare l'una con l'altra senza sovrapporsi, e che sul lungo periodo (un secolo o due) era probabile che esse coprissero il margine della zolla del Pacifico. In altri termini, l'intero margine di zolla avrebbe potuto fratturarsi.

Alla Columbia University, Lynn R.

Sykes, John Kelleher e William R. McCann applicarono queste previsioni a uno studio sismico lungo i principali margini di zolla di tutto il mondo. Molte località erano degne di nota; ciascuna era stata sede di un grande terremoto più di un secolo prima. Da allora, però, non vi era stato in quel sito alcun terremoto che avesse rappresentato una frattura significativa del margine di zolla. Sykes e collaboratori definirono ognuno di questi luoghi un «*gap* sismico»: una «lacuna sismica» candidata a un grande terremoto nei prossimi decenni. Alcune lacune sismiche si sono già fratturate, confermando le previsioni.

La faglia di San Andreas, nella California meridionale, è un'importante lacuna sismica. Ciò non significa che la faglia reagisca nello stesso modo per tutta la sua lunghezza alla dislocazione delle zolle crostali, né che l'identificazione di lacune sismiche sia l'unica base per la previsione a lungo termine dei terremoti. Su un segmento della faglia di San Andreas, nella California centrale, la dislocazione è caratterizzata da un lento scorrimento asismico. In altri luoghi la faglia libera la tensione elastica accumulata in piccoli «salti» caratteristici, non più lunghi di qualche centimetro. In altri luoghi ancora le dislocazioni sono tipicamente comprese fra sette e 10 metri. Episodi di scorrimento di quest'ultimo tipo sono stati la causa dei grandi terremoti del 1857 e del 1906.

Mediata su un periodo di tempo sufficientemente lungo, la somma dei vari scorrimenti o dislocazioni - ossia il lento scorrimento asismico, la dislocazione di faglia che accompagna i terremoti e le deformazioni anelastiche come il ripiegamento crostale - deve essere uguale

È possibile oggi spiegare il ciclo sismico nel contesto della teoria della tettonica a zolle. Fondamentalmente un terremoto è la conseguenza di un'improvvisa frattura lungo la faglia, là dove due grandi zolle crostali stanno scorrendo l'una rispetto all'altra. Prima della frattura, le zolle si muovono con regolarità a profondità superiori a 10-12 chilometri, dove la crosta è sufficientemente calda da essere plastica (*frecce in colore*), mentre a profondità minori, dove la crosta è fredda e quindi fragile, si accumula tensione in corrispondenza di «asperità», ossia di punti bloccati (*in nero*). Nel momento della frattura (*a*) un'asperità si rompe, cosicché lo scorrimento della crosta (*in colore intenso*) si diffonde attraverso la parte superiore, bloccata, della faglia. In cento secondi circa lo scorrimento rapido si arresta (*b*) e nel corso di alcuni mesi si hanno le repliche, o scosse successive (*pallini in nero, c*). Può anche aver luogo un ulteriore scorrimento in superficie o in prossimità della superficie. Molto tempo dopo, per esempio a distanza di 50 o 100 anni, può darsi che la parte adiacente della faglia si rompa dando origine a un grande terremoto (*d*). Infine, probabilmente dopo due secoli, il ciclo ricomincia: la dislocazione della parte profonda della crosta sottopone a una nuova tensione la parte superiore, fragile, della crosta (*e*).

Lo scavo esplorativo sul Pallet Creek, lungo la faglia di San Andreas, 55 chilometri circa a nord di Los Angeles, effettuato da Kerry E. Sieh del California Institute of Technology ha rivelato profondi sconvolgimenti in strati di sabbia (*puntini*), di ghiaia (*punteggiatura irregolare*) e torba (*in nero*), indicativi di episodi successivi di fagliatura associati ad antichi terremoti; l'età degli strati di torba, determinata con il carbonio radioattivo, delimita il momento dei sismi. Lo strato di torba superiore è ininterrotto ed è al di sopra delle faglie più antiche. Altre conseguenze dei terremoti sono il ripiegamento e la deformazione dei sedimenti a quel tempo soffici. In alcuni luoghi le caratteristiche del materiale sabbioso lungo una linea di faglia indicano che il sisma ha provocato la liquefazione della sabbia. Le ricerche paleosismologiche di Sieh sono utili alla previsione dei terremoti poiché forniscono statistiche sismiche per un periodo molto più lungo di quello storico.

allo spostamento fra due zolle. Ciò conduce nel modo più naturale all'idea della possibilità di una previsione sismica a lungo termine sulla base di quello che potrebbe essere designato come un bilancio degli scorrimenti. Nel 1970 uno di noi (Wallace) applicò la teoria del riassestamento elastico formulata da Reid unitamente al tasso di scorrimento dedotto dalla dislocazione delle strutture geologiche trasversalmente alla faglia di San Andreas negli ultimi 20 milioni di anni. Il risultato fu la previsione a lunghissimo termine, chiamata più propriamente determinazione del potenziale sismico. In base a essa, in qualche punto, lungo la faglia, devono aver luogo grandi terremoti a intervalli compresi fra 50 e 200 anni.

In anni recenti si è reso disponibile un altro mezzo per la previsione a lungo termine: nuove tecniche geologiche hanno consentito l'affermarsi della paleosismologia. Queste tecniche combinano studi morfologici dettagliati del terreno nelle zone di faglia, l'analisi di strati sedimentari deformati nelle pareti di fosse scavate attraverso faglie attive e la determinazione dell'età del materiale carbonioso trovato nei sedimenti per mezzo della radiometria, ossia la misurazione delle concentrazioni di isotopi del carbonio. In questo modo è stato possibile determinare il tempo approssimativo di ricorrenza e l'energia relativa dei grandi terremoti dell'antichità. Queste tecniche rendono quindi disponibile una «linea di base» degli eventi sismici più lunga rispetto a quella fornita da qualsiasi documentazione storica.

La linea di base più completa stabilita finora è stata ottenuta nel Pallet Creek, sulla faglia di San Andreas, 55 chilometri circa a nord di Los Angeles, da Kerry E. Sieh, del California Institute of Technology. Nel Pallet Creek episodi successivi di scorrimento lungo la faglia di San Andreas hanno sconvolto ripetutamente gli strati sedimentari del letto del torrente bloccando nello stesso tempo il drenaggio, cosicché gli strati frantumati sono stati ricoperti da nuovi sedimenti. In breve, l'interruzione del drenaggio e ripetuti sconvolgimenti hanno dato origine a uno schema complesso di strati fratturati e rifratturati di torba, silt e sabbia. Verso la metà degli anni settanta Sieh vi eseguì scavi analizzando ogni dislocazione. Dagli strati di torba venne prelevato il materiale per eseguire la datazione con il carbonio radioattivo e fu possibile quindi dedurre l'età di ogni episodio di fagliatura.

L'analisi iniziale di Sieh rivelò la storia di almeno nove grandi terremoti avvenuti negli ultimi 1400 anni. Poi con il proseguire degli scavi, il numero dei terremoti salì a 12. Oggi Sieh pensa che l'intervallo fra grandi terremoti lungo il segmento di Mojave della faglia di San Andreas - il segmento su cui si trova il Pallet Creek - sia compreso fra 50 e 300 anni. Secondo Sieh, l'intervallo medio di ricorrenza è di 140-150 anni.

Questo valore dovrebbe essere confrontato con la storia geologica recente del sito. L'ultimo grande terremoto nel segmento di Mojave della faglia di San Andreas si è avuto nel 1857. Gli anni trascorsi sono quindi 128, un numero vicino in maniera preoccupante al valore calcolato da Sieh per l'intervallo medio fra due grandi terremoti. Inoltre la velocità di deformazione lungo la faglia di San Andreas è di 3,5 centimetri all'anno e quindi la deformazione accumulatasi a partire dal 1857 - un totale di 4,5 metri - potrebbe generare in qualsiasi momento un terremoto di magnitudo 7,5 o più. (Un terremoto di magnitudo 5 o più è considerato potenzialmente dannoso. Il passaggio da una magnitudo alla successiva - per esempio da 5 a 6 - rappresenta un aumento di 10 volte nel movimento tellurico e un aumento di circa 30 volte nell'energia del sisma.)

Il valore trovato da Sieh per l'intervallo medio di ricorrenza è stato incorporato nella previsione preparata dal Geological Survey per il National Security Council. In quell'analisi si stima che la

probabilità di un terremoto di magnitudo 8,3 lungo la parte meridionale della faglia di San Andreas sia fra il 2 e il 5 per cento all'anno, ossia si aggiri attorno al 50 per cento per i prossimi 20-30 anni.

Nella California settentrionale non è stato ancora trovato lungo la faglia di San Andreas un sito paleosismico che fornisca una quantità d'informazione paragonabile a quella offerta dal sito del Pallet Creek. Ciononostante i dati disponibili fanno pensare a un intervallo di ricorrenza compreso fra 150 e 300 anni per un terremoto di magnitudo superiore a 7,5 lungo quella parte della faglia di San Andreas che si fratturò l'ultima volta nel terremoto del 1906. Alla catastrofe del 1906 sono seguiti cinquant'anni di quiescenza sismica, ma dopo il 1957 si sono avuti terremoti di grado moderato, come nei decenni anteriori al 1906. William L. Ellsworth e collaboratori del Geological Survey suggeriscono perciò che per i prossimi quattro decenni la prospettiva più probabile sia quella di terremoti di magnitudo compresa fra 6,5 e 7, i quali potrebbero provocare localmente gravi danni.

I processi sul lungo termine di accumulo e di liberazione di tensione sono abbastanza ben compresi; sono invece ancora in gran parte oscuri i processi-chiave per la previsione sismica a breve termine. Per esempio, non si sa con precisione come il movimento della crosta plastica inferiore crei tensioni nella crosta fragile sovrastante. Fatto forse più importante, non conosciamo ancora nei particolari in che modo la crosta fragile ceda e si rompa. In quale misura, per esempio, è uniforme la velocità di accumulo della tensione nella parte fragile della crosta? Quale significato va attribuito a variazioni di velocità? Che cosa esattamente dà inizio a un terremoto? È una sollecitazione esterna a innescarlo o un improvviso balzo nell'accumulo di tensione? Oppure si tratta di un processo progressivo di cedimento delle rocce crostali o della zona stessa di faglia?

Al centro di tutti gli sforzi diretti alla previsione a breve termine si colloca una strategia particolare. Nel periodo che precede la scossa sismica principale vanno osservati alcuni fenomeni geologici. Alcuni risultati sono già disponibili. Prima di molti grandi terremoti la sismicità di fondo della regione sembra aumentare, gradualmente o in modo brusco. Gli esempi più chiari sono stati forse gli sciami di piccoli sismi che precedettero il terremoto del 1975 a Haicheng, nella provincia di Liaoning in Cina. Gli sciami cominciarono circa due mesi prima della scossa principale, ma i ricercatori cinesi avevano già concluso che la regione aveva il potenziale per un terremoto molto forte. Essi avevano fondato la loro previsione su uno schema a lungo termine di migrazione sismica: sembrava che le posizioni dei sismi si spostassero lungo una fascia lunga varie centinaia di chilometri. Inoltre alcune misurazioni geodetiche (ossia, letteralmente, misurazioni della morfologia del terreno) avevano suggerito che si stavano accumulando deformazioni.

Nel sito stesso non si erano avuti nei decenni precedenti alla scossa principale se non alcuni terremoti di entità molto modesta. Quando però, nel febbraio 1975, gli sciami sismici si fecero intensi, gli abitanti che vivevano in case in muratura non rinforzate furono evacuati, cosicché la sera del 4 febbraio, quando ebbe luogo la scossa principale e molte case crollarono, fu evitato senza dubbio un gran numero di perdite umane. Di contro il terremoto di Tangshan, avvenuto circa 18 mesi dopo, non fu preceduto da alcuna scossa precorritrice e provocò un numero di vittime stimato ufficialmente in 240 000.

La sismicità è un indizio importante di grandi terremoti, ma non è l'unico segno premonitore. A volte, anche se non sempre, prima di grandi terremoti si scopre una deformazione accelerata della crosta. Questo fenomeno fu rilevato per esempio prima del terremoto del 1966 avvenuto a Parkfield, lungo la faglia di San Andreas. Alcuni esperimenti di laboratorio e certi modelli teorici del processo sismico suggeriscono che una deformazione accelerata - chiamata scorrimento presismico - sia intrinseca ai terremoti. Si pensa che il cedimento delle rocce della crosta sia preceduto dallo sviluppo di piccole crepe o punti deboli, che poi crescono dando origine a fratture catastrofiche. L'identificazione dei precursori del sisma all'interno del contesto della deformazione crostale, che presenta continue fluttuazioni, rimane uno degli obiettivi principali delle ricerche sulla previsione dei terremoti.

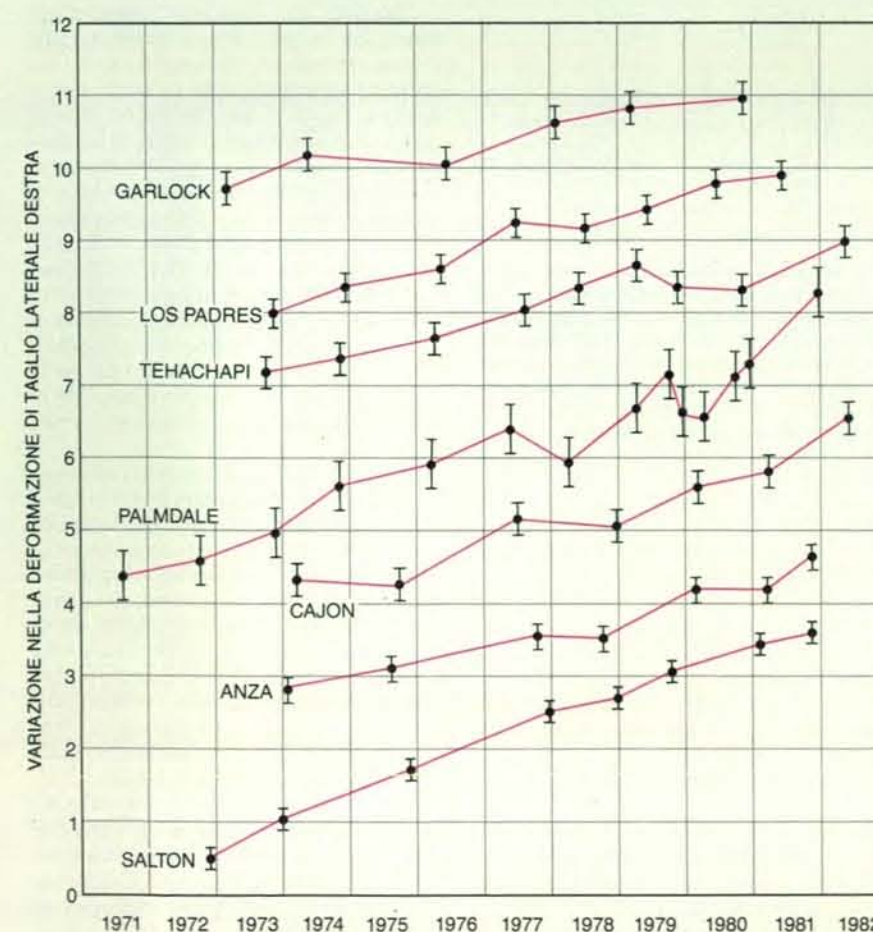
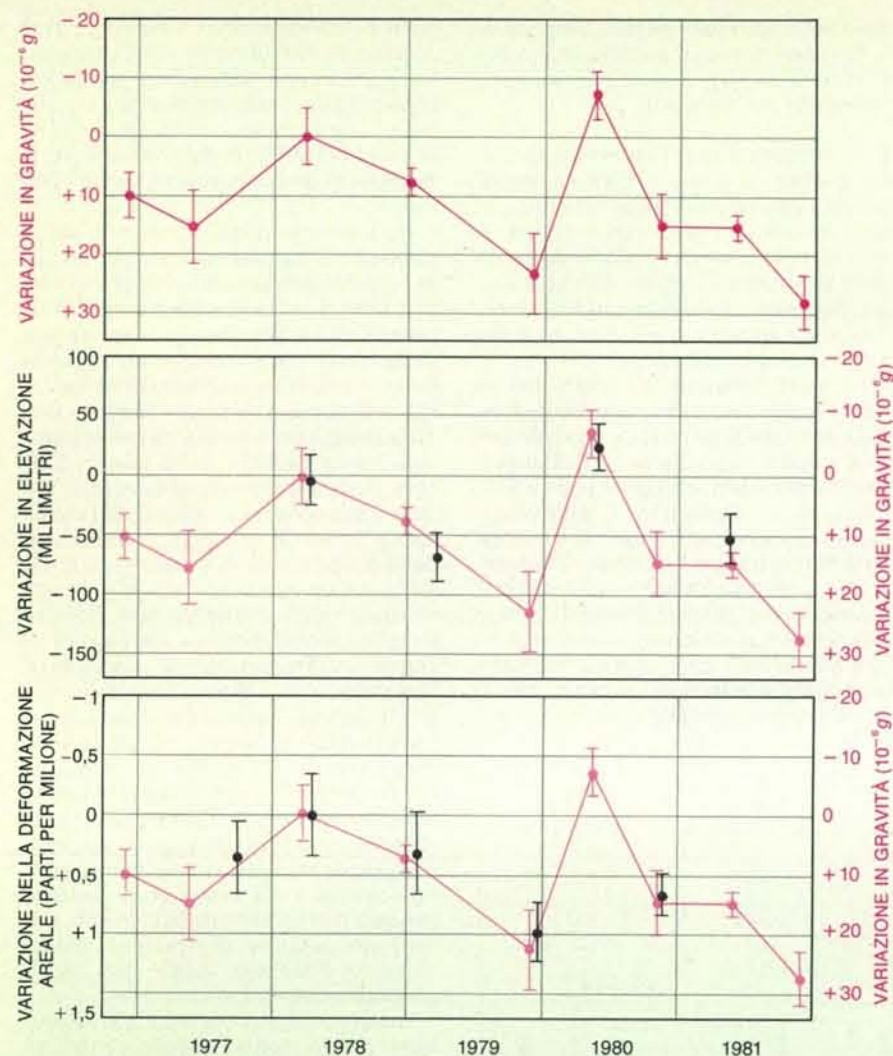
La strumentazione richiesta per controllare la crosta e rilevare eventi anomali che potrebbero preannunciare un terremoto ha fatto rapidi progressi nel decennio scorso, per quanto riguarda sia la capacità sia l'attendibilità dei singoli dispositivi. Gli strumenti sono inoltre aumentati numericamente, coprendo un'area sempre più estesa.

La parte principale di questo sforzo ricade sugli economici e attendibili gruppi di sismometri pronti per l'installazione e messi a punto da Jerry P. Eaton e collaboratori del Geological Survey. Costituiti da un sismometro e dall'elettronica di supporto, ve ne sono in funzione oltre 500 in tutta la California. Per mezzo di collegamenti telefonici o radio essi trasmettono di continuo segnali sismici ai centri di registrazione e di elaborazione a Menlo Park e a Pasadena. Nel 1984 sono stati registrati in tal modo più di 10 000 terremoti. Per mezzo di hardware e software messi a punto da Rex Allen e James O. Ellis del Geological Survey vengono determinate automaticamente posizione, profondità e magnitudo di ogni evento sismico; vi sono giorni in cui i calcolatori elaborano automaticamente i dati concernenti anche 1200 scosse.

La strumentazione comprende anche varie reti di dispositivi laser per il controllo della lunghezza dei lati di una rete di triangoli. Una variazione nella lunghezza di un lato fornisce una misura dello sforzo cui è sottoposta la crosta lungo quella linea; una variazione dell'area delimitata da ciascun triangolo fornisce in aggiunta una misura volumetrica della deformazione della crosta. Con correzioni appropriate per tener conto delle caratteristiche ottiche dell'atmosfera, ottenute mediante misurazioni della temperatura dell'aria lungo la visuale da un aereo in volo è possibile ottenere valori con un margine d'errore inferiore al centimetro su una linea di 30 chilometri. Recentemente sono entrati in servizio alcuni sistemi laser a due colori. I fasci di diverso colore hanno proprietà ottiche differenti, rendendo in tal modo possibile il calcolo di variazioni atmosferiche direttamente dal tempo di trasmissione dei fasci. In tal modo le variazioni atmosferiche vengono automaticamente approssimate. La deformazione verticale della crosta corrisponde a variazioni di elevazione del terreno, che vengono misurate direttamente determinando l'intensità locale del campo gravitazionale terrestre.

Andrebbe detto che tutte queste tecniche danno risultati soddisfacenti solo attraverso rilevamenti ripetuti; di conseguenza comportano un forte dispendio di tempo e di denaro e sono rappresentative solo di una media regionale e temporale in un dato istante. Per questo motivo negli ultimi 15 anni si è cercato in ogni modo di mettere a punto strumenti attendibili che misurino con continuità la deformazione in un punto preciso della Terra. Questo sforzo ha permesso di stabilire che la crosta terrestre entro alcuni metri dalla superficie è così «rumorosa», ossia è fonte di tanti dati spuri, da mascherare i segnali di deformazione significativi. Fonti di rumore sono le variazioni diurne e stagionali di temperatura, le precipitazioni, le fluttuazioni nella pressione barometrica e quelle nel livello della falda freatica. La speranza quindi di poter mettere a punto uno strumento economico di vasta applicazione, alla stregua di un sismografo, è rimasta per ora irrealizzata.

Uno strumento in grado di effettuare misurazioni continue della deformazione è un rilevatore di scorrimento asismico formato da un filo metallico, lungo per esempio 30 metri, teso all'interno di un tubo situato alla profondità di alcuni metri con un misuratore di deformazione a una estremità. Un altro strumento promettente è un estensimetro da pozzo

Le reti di laser per la determinazione delle distanze formano un insieme di triangoli che attraversano la faglia di San Andreas nella California meridionale. In ciascuna rete i tempi di percorrenza dei fasci laser forniscono una misura della distanza su linee di base di anche 30 chilometri consentendo ai ricercatori di dedurre dalle variazioni delle distanze l'accumularsi di tensione nella crosta, indizio di un probabile terremoto. Le misurazioni della deformazione di taglio laterale destra - la distorsione elastica che tende a causare dislocazione sulla faglia di San Andreas - illustrate nel grafico a fianco sono state eseguite da James C. Savage e collaboratori del Geological Survey. Tutti i valori tendono a salire indicando l'accumularsi della deformazione. Gli scostamenti da un aumento costante potrebbero precorrere un terremoto. I materiali della crosta non possono resistere a una deformazione di taglio superiore a 10^{-4} radianti circa (un radiante è un angolo di circa 57 gradi) e quindi una deformazione di taglio che si accumuli nella crosta a una velocità di 5×10^{-7} radianti per anno implica la ricorrenza di un sisma ogni 200 anni circa.

A Tejon Pass, nella California meridionale, gravità, elevazione e deformazione orizzontale della crosta hanno tutte presentato variazioni. Nei grafici sono tracciati l'intensità del campo gravitazionale locale (*in alto*), l'altezza del terreno (*al centro*) e la deformazione della crosta, come risulta dalla variazione dell'area superficiale (*in basso*). Le correlazioni fra i tre parametri inducono a pensare che ciascuna di queste misurazioni indipendenti rifletta fluttuazioni reali della crosta e non un semplice «rumore» dovuto al sistema di monitoraggio locale. Parametri come questi vengono studiati come possibili segnali premonitori di terremoti. I tre insiemi di misurazioni sono stati eseguiti da Robert Jachens e collaboratori del Geological Survey.

che è stato progettato da I. Selwyn Sacks e collaboratori della Carnegie Institution of Washington. Esso è formato essenzialmente da un pallone pieno di fluido e forzato sul fondo di un pozzo profondo. La quantità di fluido estruso dal pallone in risposta alla compressione o all'espansione della crosta è misurata continuamente per mezzo di dispositivi elettronici. Molti strumenti di questo tipo sono attualmente installati in California. L'osservazione del livello dell'acqua nei pozzi e di fenomeni geochimici come la concentrazione di rado nell'acqua del pozzo è nel migliore dei casi ambigua. I rilevamenti geodetici eseguiti da satelliti stanno migliorando; in futuro potranno offrire misurazioni del movimento dei blocchi crostali abbastanza esatte da poter essere utilizzate nella previsione dei terremoti.

Che cosa ci rivelano dunque i dati disponibili sulla geologia in tempo reale della California? In che modo contribuiscono alla previsione sismica a breve termine? L'anno 1980 può essere considerato un riferimento importante. Fra le variazioni geodetiche osservate prima del 1980 una fu particolarmente rilevante: una parte della California meridionale stava sollevandosi. Al Geological Survey, Robert O. Castle e collaboratori passarono in rassegna i dati di decenni tratti dall'US National Geodetic Survey e da altre fonti. Essi giunsero alla conclusione che, dal 1960 al 1962, la regione attorno a Palmdale, una cinquantina di chilometri a nord di Los Angeles, si era sollevata di circa 25 centimetri. Dal 1972 al 1974 si ebbe un altro sollevamento di 15 centimetri. Il massimo del sollevamento si spostò inoltre di 125 chilometri verso est-sud-est e l'area del sollevamento si estese verso sud-est, raggiungendo quasi il Salton Sea. La dislocazione di un volume di crosta così cospicuo suggeriva che potesse essere imminente un grande terremoto.

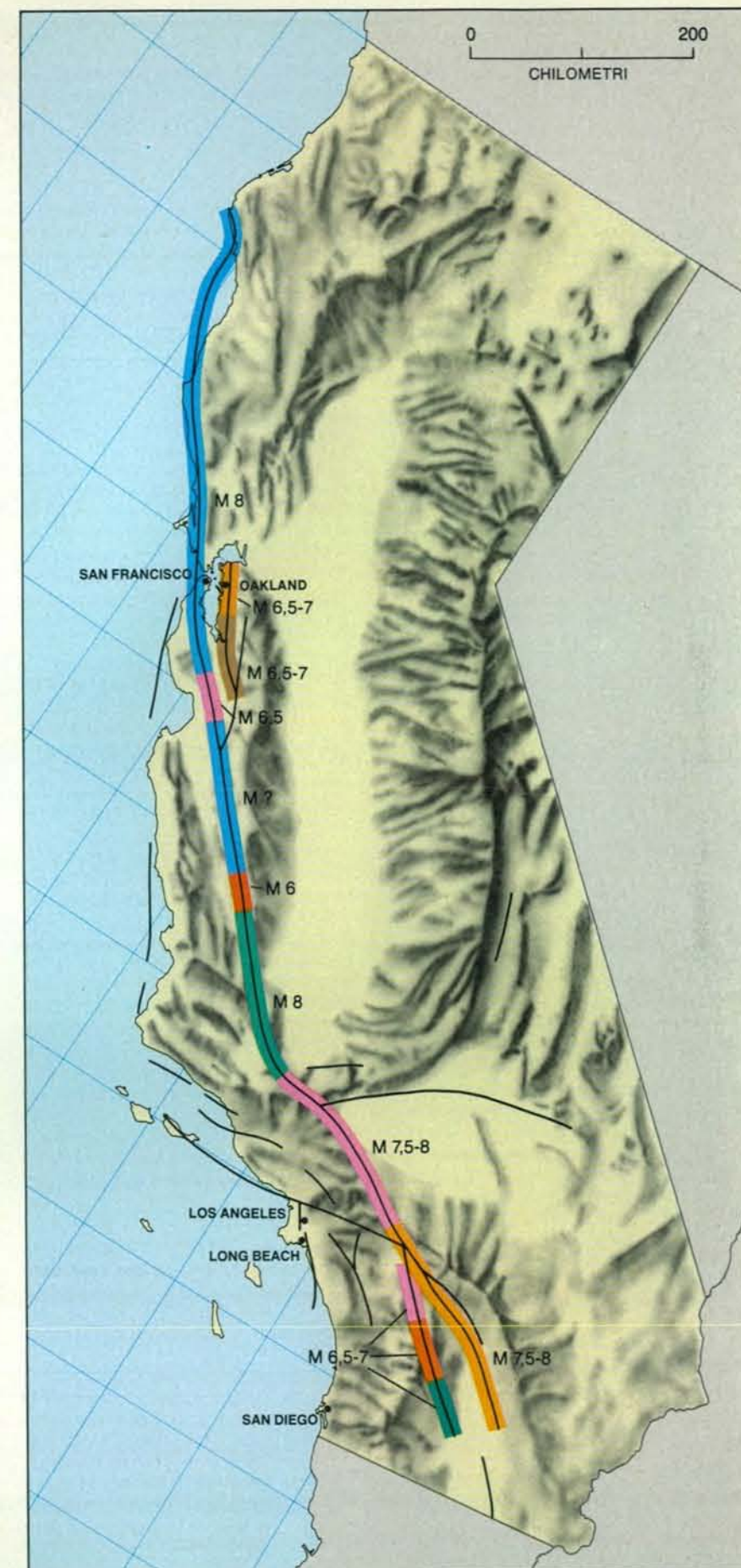
Altrove nel mondo - per esempio a Niigata in Giappone - i terremoti erano stati preceduti dalla deformazione locale della crosta. Inoltre il sollevamento aveva luogo in una regione ellittica sovrastante gran parte della faglia di San Andreas, la cui frattura aveva prodotto il grande terremoto del 1857. Certamente alcuni dei dati altimetrici più antichi di cui ci si era serviti nella valutazione del sollevamento di Palmdale sono discutibili. Di conseguenza sono in corso nuovi rilevamenti per ridurre al minimo la possibilità di errori sistematici. Forse la prova più conclusiva del fatto che si stanno scoprendo veramente fenomeni di deformazione della crosta è stata fornita da Robert Jachens e collaboratori del Geological Survey. Essi hanno trovato una forte correlazione fra tre insiemi indipendenti di misurazioni effettuate nella California meridionale su un periodo di cinque anni: l'intensità locale del campo gravitazionale terrestre, la variazione di elevazione e quella di deformazione.

Ulteriori dati geodetici suggerivano che erano in corso altri mutamenti della crosta, mutamenti che tendevano ad accrescere la potenzialità di scorrimento lungo la faglia di San Andreas. Rilevamenti di deformazione orizzontale eseguiti da James C. Savage e collaboratori del Geological Survey su un periodo di 10 anni hanno rivelato sul lungo termine un accorciamento della crosta in direzione nord-sud. Poi, nel 1978 o 1979 l'andamento della deformazione si è alterato in modo tale da accrescere lo sforzo di taglio lungo la direzione della faglia di San Andreas. Lo scorrimento lungo la faglia risultava dunque più probabile di prima, senza tuttavia conoscerne il valore di soglia.

Nel 1979 anche i dati relativi alla concentrazione di rado in due pozzi lungo il margine settentrionale del bacino di Los Angeles hanno presentato aumenti molto marcati. I dati disponibili coprivano un periodo di un solo anno; ciononostante l'aumento è stato giudicato significativo. Si sapeva, infatti, che in Unione Sovietica e nella Repubblica Popolare Cinese aumenti del genere avevano preceduto numerosi grandi terremoti. Presumibilmente il gas veniva liberato dalla deformazione delle rocce crostali. Considerati congiuntamente, i dati sulla ricorrenza sul lungo termine e le variazioni geofisiche osservate hanno suscitato nel 1980 il timore che nella California meridionale un grande terremoto potesse essere imminente.

Da allora, però, il tentativo di formulare una previsione specifica per la California meridionale è divenuto piuttosto problematico. Sono stati individuati vari siti di potenziale sismico. Oggi si tende ad assegnare una maggiore probabilità a terremoti più frequenti, ma meno intensi di quello del 1857. Certi segmenti della faglia di San Andreas non si sono mai fratturati in tempi storici; la tensione che si è accumulata lungo quei tratti potrebbe farne i siti più probabili per il prossimo grande terremoto. (Alcuni ricercatori sostengono che un sito molto probabile è la parte della faglia di San Andreas a sud-est di San Bernardino.) Infine, in California vi sono altre grandi faglie oltre a quella di San Andreas. Anch'esse potrebbero generare grandi terremoti, e di fatto pare che sia così, anche se con intervalli di ricorrenza medi di oltre 1000 anni. Alcune di queste faglie sono esposte in superficie; altre sono sepolte. Una faglia profonda circa 10 chilometri ha prodotto il terremoto di moderata intensità, ma dannoso, che colpì Coalinga nel 1983.

Un importante esperimento per la previsione dei terremoti è stato di recente ideato da Allan G. Lindh e collaboratori del Geological Survey lungo il segmento di Parkfield della faglia di San Andreas, nella California centrale. Vi si sono avuti terremoti di magnitudo compresa fra 5,5 e 6 una volta ogni 22 anni circa, il più recente nel 1966. (L'unica eccezione in più di un secolo è stato il terremoto del 1934, che avvenne con 10 anni di anticipo.) Oggi è stata installata in quest'area una schiera di strumenti, compresi rilevatori di scorrimento asismico, estensimetri da pozzo, dispositivi laser a due colori per la misurazione delle distanze. Si spera in questo modo di apprendere esattamente quali variazioni geofisiche e quali segni premonitori

La probabilità che abbia luogo un terremoto in un qualche momento nei prossimi 30 anni lungo segmenti della faglia di San Andreas è stata stimata da Allan G. Lindh e collaboratori del Geological Survey. Ogni segmento ha una probabilità diversa in conseguenza della diversa situazione geologica. Alcuni segmenti non seguono il movimento della zolla del Pacifico rispetto alla zolla nordamericana; lungo quei segmenti il pericolo di una liberazione di tensione da parte di un grande terremoto è presumibilmente massimo. Altri segmenti presentano scorrimento asismico, mentre altri ancora liberano tensione in modesti sismi. Tutte le previsioni sismiche sono concordi nell'associare i terremoti alla frattura di un dato segmento di faglia. Ciascun terremoto avrebbe una magnitudo (*M*) caratteristica.

avranno luogo prima del terremoto atteso per il 1988, con un'approssimazione di qualche anno in più o in meno. Nel 1980 William H. Bakun del Geological Survey previde con esattezza luogo e intensità del terremoto di magnitudo 6 che si ebbe nei pressi di Morgan Hill, in California, nell'aprile dello scorso anno. In quell'occasione non venne rilevato alcun segno premonitore che potesse consentire una previsione a breve termine del momento del terremoto.

Come si possono sfruttare nel modo migliore a beneficio della società le nuove conoscenze su come e quando potrebbero aver luogo terremoti di vaste proporzioni? Fra le strategie a lungo termine, la prima linea di difesa consiste nell'includere, nella progettazione e nella costruzione degli edifici, tutti i materiali e gli accorgimenti in grado di conferire loro una resistenza antisismica. In California i regolamenti sull'edilizia sono di competenza locale, ma si basano su raccomandazioni fatte da un ente statale, la Structural Engineers Association. Le raccomandazioni si basano sulle analisi tecniche e anche sull'esperienza acquisita empiricamente osservando il modo in cui si sono comportati effettivamente edifici di vario tipo in occasione di terremoti. Così i regolamenti edilizi incorporano una serie di requisiti minimi. Inoltre la legislazione statale impone condizioni speciali relativamente a scuole, ospedali e dighe e richiede studi particolari per gli edifici che devono essere ubicati in prossimità di faglie attive. In generale ricevono una particolare attenzione progetti edilizi importanti, come i grattacieli di Los Angeles e di San Francisco, e progetti speciali, come i reattori nucleari. D'altro canto la maggior parte degli edifici unifamiliari con struttura in legno sono considerati intrinsecamente antisismici.

Non va trascurato il fatto che le mode architettoniche mutano, e con esse i materiali e le tecniche costruttive disponibili. Perciò accade che in un momento qualsiasi molte delle costruzioni esistenti non siano conformi alla maggior parte dei regolamenti edilizi recenti o allo stato attuale della scienza delle costruzioni. Spesso accade inoltre che edifici di dimensioni intermedie vengano costruiti rispettando solo le prescrizioni di legge minime, senza prendere opportune misure per conferire agli edifici una speciale resistenza alle scosse sismiche. Nella nostra società questi edifici rappresentano una grande parte dell'inventario totale delle strutture. Inoltre, ancora, la cura con cui un progetto edilizio viene tradotto in realtà - il controllo qualitativo e l'attenzione ai particolari, per esempio nella messa in opera di strutture di rinforzo in acciaio e nell'esecuzione di saldature - è cruciale per la buona riuscita di una costruzione. A volte tale cura viene sacrificata ai fini di accelerare la costruzione o per ridurre i costi.

In California molte comunità stanno realizzando piani per un uso del territorio che riduca al minimo i potenziali rischi sismici. È chiaramente imprudente costruire strutture destinate all'uomo a cavallo di faglie o su terreni franosi che potrebbero spostarsi catastroficamente in caso di terremoto. La destinazione di queste aree a parchi, strade e altri spazi aperti creerebbe invece ben pochi rischi per la vita e la sicurezza.

Previsioni sismiche a breve termine potrebbero accrescere il margine di sicurezza, in particolare per persone che vivono o lavorano in edifici non conformi alle norme antisismiche correnti o per persone che si trovano in prossimità di strutture il cui crollo potrebbe rappresentare un pericolo. Alcune semplici strategie attuabili con una previsione a breve termine sono l'evacuazione di strutture rischiose - non di intere città - per alcune ore o per alcuni giorni, la dislocazione all'aperto di ambulanze e di automezzi dei vigili del fuoco e la messa in stato di allerta del personale addetto a certi compiti specifici. In un impianto nucleare precauzioni aggiuntive potrebbero andare da semplici esercitazioni e da un riesame delle misure da adottare in caso di disastro, ad azioni più dispendiose, come la riduzione della potenza o addirittura la chiusura temporanea dell'impianto. A condizioni analoghe devono far fronte molti impianti industriali. La semplice rimozione di apparecchiature fragili - per esempio strumenti elettronici e prodotti chimici e farmaceutici - dagli scaffali potrebbe già contribuire a ridurre vistosamente i danni.

A seguito della legislazione statale e di finanziamenti sia statali sia federali, è stato istituito nel 1980 il Southern California Earthquake Preparedness Project o SCEPP. Il progetto riguarda lo sviluppo di piani per sfruttare efficacemente le previsioni, allo scopo di ridurre il danno sociale ed economico di un terremoto, di far fronte a terremoti non previsti e di organizzare successivamente la ripresa. Nel frattempo sono in corso di messa a punto meccanismi amministrativi per rendere pubbliche le informazioni sulla probabilità di terremoti in California. A livello federale la responsabilità di pubblicazioni sui rischi geologici, comprese le previsioni sismiche è di competenza del direttore dell'US Geological Survey, che lavora a stretto contatto con la sua controparte statale, il capo della California Division of Mines and Geology. Gruppi di specialisti in scienze della Terra a livello sia statale sia federale daranno la loro consulenza nel giudicare la validità scientifica di una previsione.

L'adozione di misure in risposta alle informazioni ricevute (o a un terremoto) è di competenza della Federal Emergency Management Agency e dello State Office of Emergency Services. Fino a oggi il Geological Survey ha reso noti due «pareri» (*watch*) sismici ufficiali, uno per un grande terremoto nella California meridionale e uno per un terremoto di moderata intensità, ma potenzialmente dannoso, nella regione dei Mammoth Lakes della California orientale. Una serie di terremoti di magnitudo compresa fra 5 e 6 seguì effettivamente al secondo «parere». In vari altri casi anomalie geofisiche e geologiche che sembravano preannunciare un terremoto imminente sono state esaminate a livello sia statale sia federale, ma non sono state ritenute sufficienti per decretare lo stato di allarme.

Gli studiosi di scienze della Terra stanno acquisendo fiducia nella capacità di fare previsioni a lungo termine sulla probabilità di un terremoto. I prossimi passi avanti consisteranno nello sviluppo di sistemi integrati di misurazione, raccolta e analisi rapide e automatizzate di dati, e nell'applicazione delle analisi alle complessità della deformazione crostale. La conversione delle previsioni a lungo termine in previsioni a breve termine e quindi in allarmi porranno problemi di difficile soluzione. Per esempio, i dati relativi al lungo termine inducono a pensare che nella California meridionale possa essere imminente un grande terremoto. Ciononostante le variazioni di breve termine nei parametri geologici osservate attualmente nella California meridionale non sono chiare. Queste variazioni sono precorritrici della scossa o sono fluttuazioni normali della crosta? Non abbiamo ancora sufficiente esperienza per dirlo. Molti esperti di scienze della Terra sono oltremodo preoccupati su come rendere nota al pubblico la loro apprensione senza suscitare allarmi ingiustificati. Nella Repubblica Popolare Cinese, dove alcune previsioni sismiche si sono rivelate esatte, ve ne sono state molte di infondate. Se alcune variazioni osservate oggi in California fossero riscontrate in Cina sarebbero sufficienti per una previsione a breve termine.

Che cosa faranno gli esperti di scienze della Terra quando si troveranno di fronte simultaneamente a tre o quattro tipi diversi di anomalie? Potrebbe trattarsi di una sequenza di scosse piccole o moderate lungo un segmento in precedenza quiescente della faglia di San Andreas o l'inizio di uno scorrimento asismico di faglia in un momento in cui l'accumulo della deformazione orizzontale sta accelerando. Può darsi che i ricercatori siano comprensibilmente riluttanti a emanare un allarme: il costo di un falso allarme è infatti potenzialmente grande. È probabile che anche gli abitanti siano riluttanti ad aderire a un allarme: tale fu il comportamento infatti che si riscontrò prima dell'eruzione del Mount Saint Helens, anche se alcuni segni premonitori erano molto evidenti. Le potenziali conseguenze del prossimo grande terremoto in California sarebbero però così gravi da giustificare ogni sforzo per renderne possibile la previsione. Questa è la travagliata nascita della geologia in tempo reale.

Sorgenti di impulsi gamma

La carenza di dati osservativi ha finora impedito di determinare la causa degli intensi lampi di radiazione ad alta energia che compaiono nel cielo in luoghi e tempi imprevedibili, ma non di proporne modelli plausibili

di Bradley E. Schaefer

Circa una volta al giorno da una zona completamente imprevedibile del cielo si sprigiona un impulso (*burst*) di radiazione gamma molto intensa. Gli impulsi hanno durate caratteristiche comprese tra uno e 10 secondi, anche se alcuni le hanno avute di soli 0,01 secondi e altri addirittura di 80 secondi. In questo arco di tempo l'intensità dell'impulso raggiunge la soglia di visibilità, oscilla casualmente e poi diminuisce di nuovo fin sotto la soglia. Tranne poche eccezioni non sono stati osservati più impulsi provenienti da un'unica, esattamente identica, direzione e nessun impulso è stato identificato con sicurezza con un oggetto già noto.

Quando si trovano al massimo, questi impulsi costituiscono le emissioni di raggi gamma di gran lunga più intense del cielo. In effetti, accettando alcune ipotesi ragionevoli riguardo alla distanza da cui provengono, sembra che le loro sorgenti sviluppino una potenza per unità di volume superiore a quella di qualsiasi altro oggetto dell'universo conosciuto: secondo una valutazione producono in un secondo tanta energia quanta ne emette il Sole in una settimana.

Nonostante l'enorme quantità di energia liberata in ciascuna emissione, però, su queste sorgenti si posseggono poche informazioni. Gli impulsi infatti sono imprevedibili e cadono in una regione dello spettro estremamente difficile da studiare con precisione. La scarsità di dati osservativi lascia spazio a numerose proposte di modelli: fino a oggi ne sono state avanzate quasi 40, alcune delle quali anche esotiche come quelle dei buchi neri in esplosione, dei granelli di polvere ultrarelativistici e della fissione di elementi superpesanti. Si rimane perplessi nel constatare che quasi tutti questi modelli sono compatibili con i risultati attuali dell'osservazione.

Anche con le poche informazioni disponibili, però, negli ultimi tempi gli astronomi hanno cominciato a procedere verso il raggiungimento di un accordo su quelle che devono essere le caratteristiche generali di una sorgente di impulsi gamma. Inoltre, sono in corso di realizzazione alcuni strumenti che forse risolveranno molti dei problemi più importanti sulle proprietà degli impulsi: nel prossimo futuro sarà forse possibile determinare esattamente la causa di un impulso gamma, ponendo così fine al fervore di ipotesi iniziato a partire dalla scoperta di questi fenomeni misteriosi, oltre dieci anni fa.

Le sorgenti di impulsi gamma sono state osservate per la prima volta da un gruppo di satelliti Vela del Department of Defense degli Stati Uniti. Progettati per controllare il rispetto da parte dei sovietici di un trattato che bandisce gli esperimenti nucleari nello spazio, questi satelliti erano dotati di rivelatori per registrare l'improvviso lampo di radiazione gamma che accompagna ogni esplosione nucleare. Dopo il rilevamento di diversi lampi fu chiaro che non si trattava di esperimenti atomici, ma di una classe completamente nuova di oggetti astronomici precedentemente sconosciuti. La scoperta delle sorgenti di impulsi gamma fu presto confermata da Thomas L. Cline e Upendra D. Desai del Goddard Space Flight Center della National Aeronautics and Space Administration. Poco dopo vari altri gruppi riuscirono a rintracciare impulsi gamma nei dati relativi a esperimenti precedenti, progettati per altri scopi.

Verso la fine degli anni settanta ha avuto inizio una seconda generazione di esperimenti (si veda l'articolo *Impulsi cosmici di raggi gamma* di Ian B. Strong e Ray W. Klebesadel in «Le Scienze» n. 101, gennaio 1977). Per individuare la direzione di un impulso con la massima precisione possibile è stata realizzata una rete composta da circa una decina di rivelatori a bordo di diverse sonde che dovevano muoversi nelle regioni più interne del sistema solare.

In uno di essi, l'esperimento «Konus» (dal russo per richiamare la forma a «cono» delle sonde), quattro sonde sovietiche del tipo Venera sono state equipaggiate con rivelatori estremamente sensibili. Gli autori dell'esperimento E. P. Mazets, S. V. Golenetskii e collaboratori dell'Istituto di fisica e tecnologia A. F. Joffe di Leningrado hanno utilizzato questi rivelatori per compilare un catalogo di date, posizioni, spettri e variazioni di luminosità per centinaia di impulsi diversi.

Queste osservazioni più recenti hanno permesso di determinare la posizione di alcuni impulsi, ma è difficile ottenere informazioni più specifiche su uno di essi in particolare. Data la loro brevità, e dal momento che avvengono in posizioni imprevedibili, è impossibile individuarli mediante un rivelatore molto sensibile con un campo di osservazione ristretto: per avere una possibilità significativa di riuscire a rivelarne almeno uno è, invece, necessario utilizzare uno strumento relativamente poco sensibile, ma dotato di un grande angolo di rilevamento.

Ulteriori difficoltà derivano dal fatto che gli impulsi sono formati quasi esclusivamente da radiazione gamma, che viene assorbita dall'atmosfera terrestre: è quindi necessario lanciare i rivelatori nello spazio, un procedimento in realtà costoso e complicato. Inoltre i rivelatori di raggi gamma sono di per sé meno sensibili di quelli per le radiazioni a energia inferiore.

A fronte di queste difficoltà sperimentali, quali tipi di dati potenzialmente utili per discriminare i modelli sono disponibili? Fondamentalmente, tutte le osservazioni nella regione spettrale dei raggi gamma sono costituite soltanto da «curve di luce», registrazioni della variazione di luminosità di un impulso. Dato che ogni impulso comprende radiazioni a vari livelli di energia è necessario registrare diverse curve di luce (una per ogni livello di energia) per ciascuno. Fino a oggi sono state raccolte curve di luce relative a varie centinaia di casi e spesso quelle di un certo impulso sono state registrate da più di un satellite. Questa piccola collezione di curve di luce rappresenta la quasi totalità del nostro complesso di conoscenze a proposito delle sorgenti di impulsi gamma.

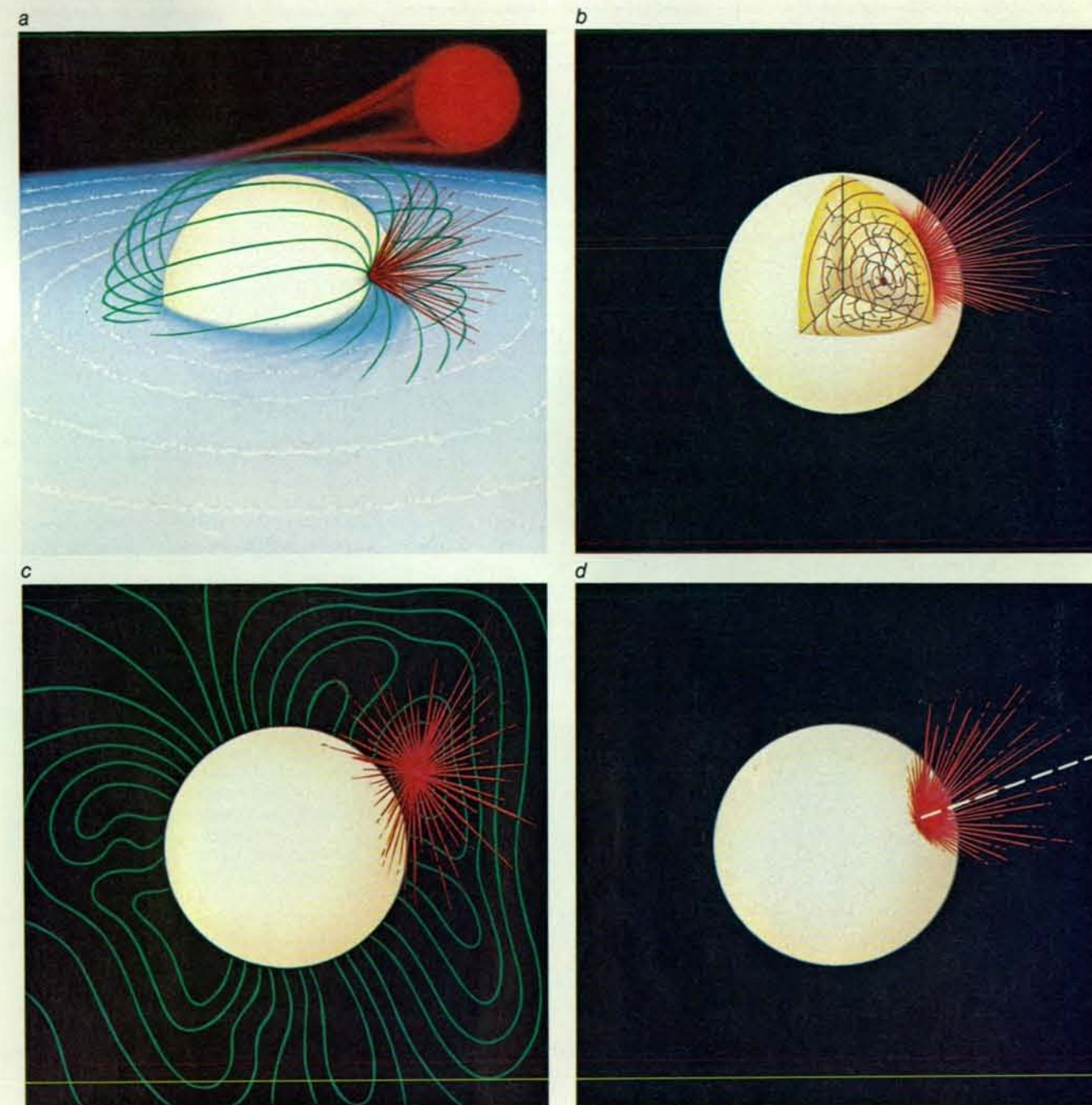
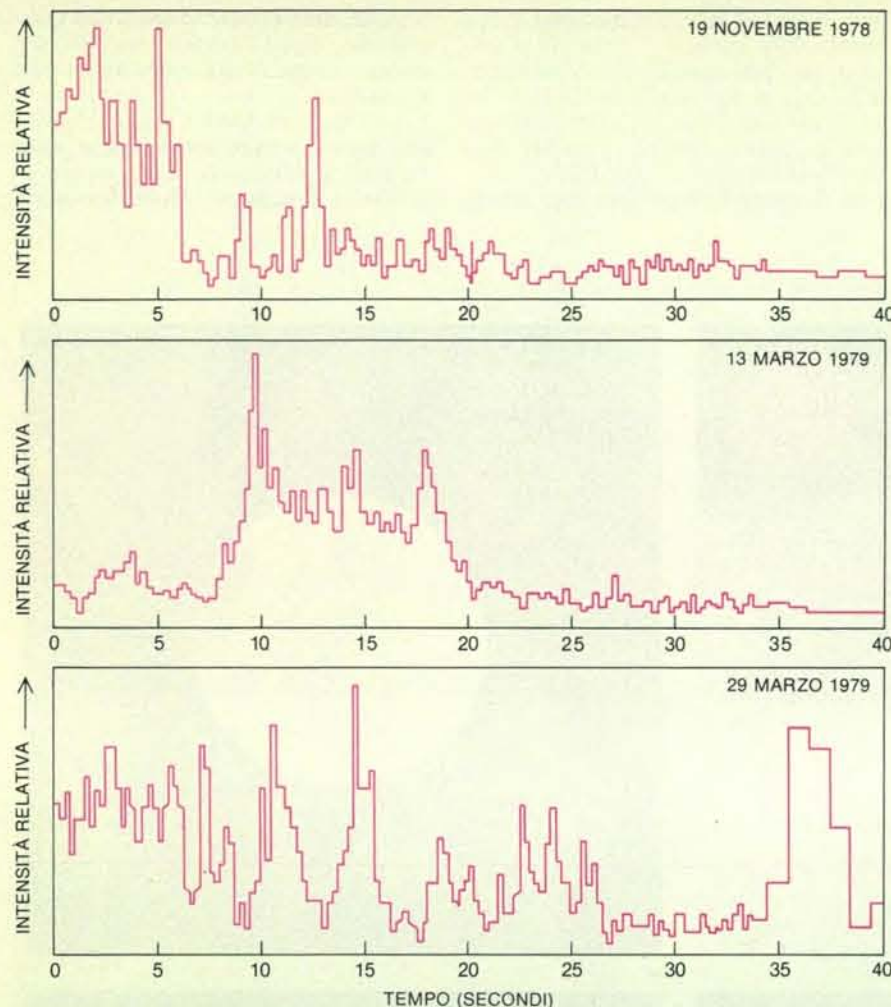
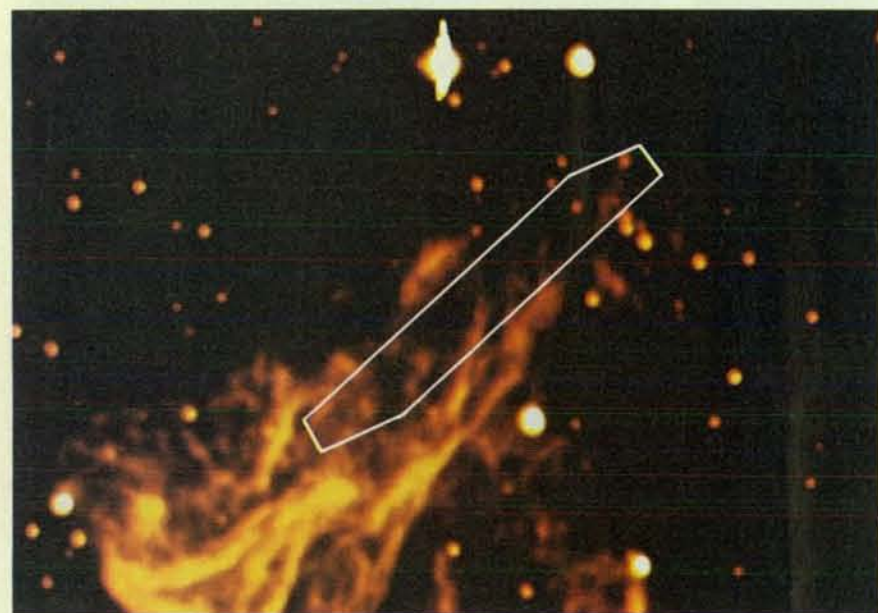
Per quanto scarsi questi dati possano sembrare, esistono sistemi per ricavare informazioni utili anche dalle curve di luce. Confrontando, per esempio, le intensità delle curve e i momenti in cui sono state registrate dai rivelatori su satelliti diversi, è possibile determinare la posizione di un impulso nel cielo con una certa precisione. Inoltre è possibile, paragonando le curve di luce relative a livelli di energia diversi per uno stesso impulso, determinare lo spettro di quest'ultimo, cioè l'intensità relativa delle diverse energie di radiazione da cui esso è costituito.

Il profilo dello spettro di una sorgente può spesso fornire informazioni sulla sua natura, ad esempio sulla temperatura o sulle dimensioni. Molti ricercatori

a

b

c

d

La penuria di dati capaci di fornire una prova definitiva ha causato una vera proliferazione di modelli delle sorgenti di impulsi gamma. I più probabili, quattro dei quali sono qui raffigurati, si basano su stelle di neutroni. In *a* parte della materia di una compagna relativamente poco luminosa *(in rosso)* si raccoglie in un disco centrato sulla stella di neutroni, per venire poi attirata lungo le linee di forza del suo campo magnetico *(in verde)* verso il polo magnetico. Accumulandosi sul polo, si comprime e si riscalda fino a esplodere emettendo raggi gamma. In *b* la stella subisce un «nucleomoto» che si propaga fino alla superficie dove le onde d'urto producono i raggi gamma. Il modello *c* attribuisce gli impulsi a irregolarità nel campo magnetico della stella: alcune linee molto irregolari possono diventare instabili formando un anello chiuso sopra la superficie della stella, che inizia a restringersi. Alla fine l'anello scompare e la notevole quantità di energia che costituiva la porzione irregolare del campo si libera in forma di radiazione elettromagnetica, parte della quale sarà fatta di raggi gamma. Gli impulsi potrebbero derivare anche dall'impatto di una cometa o di un asteroide (*d*) sulla superficie di una stella di neutroni (*traiettoria in bianco*). Data l'intensità del suo campo gravitazionale la collisione avrebbe un'energia sufficiente a produrre raggi gamma.

Le curve di luce rappresentano la variazione della luminosità di un impulso nel tempo. Queste tre, prodotte dal rivelatore del «Konus» sovietico, descrivono impulsi diversi e non presentano una struttura comune. Questa circostanza è estremamente problematica perché le curve di luce sono tra le poche informazioni osservative inequivocabili sugli impulsi gamma disponibili.

La posizione *(evidenziata nel riquadro)* dell'impulso di raggi gamma insolito che ha avuto luogo il 5 marzo 1979 è stata stabilita con notevole precisione. L'impulso è stato osservato da 11 rivelatori ben distanziati all'interno del sistema solare; la posizione dell'impulso è stata determinata confrontando in quali momenti i vari rivelatori hanno osservato il segnale.

hanno esaminato gli spettri degli impulsi che sono stati osservati, nel tentativo di individuare il processo di emissione da cui derivano. Sono noti parecchi processi radiativi che generano spettri simili a quelli osservati per gli impulsi gamma, ma nessuno riesce a spiegare in maniera soddisfacente i profili spettrali di tutti.

Per la verità sarebbe sorprendente che uno di questi meccanismi teorici riuscisse a spiegare gli spettri degli impulsi: ognuno, infatti, è valido solo in condizioni ideali, cioè con un solo valore di temperatura, di orientazione e di intensità del campo magnetico, e in un solo istante di tempo; la relazione tra la situazione reale e quelle descritte dai modelli idealizzati è probabilmente molto complessa.

Le informazioni spettrali, anche se non sono particolarmente utili per stabilire le proprietà della regione emittente, possono contenere dati relativi allo spazio che circonda quella regione. Se, per esempio, tale regione assorbisse o emettesse radiazione a una delle frequenze dell'impulso, nello spettro della sorgente osservato comparirebbe una riga. In effetti l'esperimento sovietico Konus ha individuato proprio alcune righe di questo tipo, per lo più nella banda compresa tra 40 e 70 KeV (migliaia di elettronvolt), nel 15 per cento circa degli impulsi che ha rivelato. Anche Geoffrey J. Hueter dell'Università della California a San Diego ha trovato indicazioni indipendenti sull'esistenza di queste righe nei dati raccolti dal satellite HEAO-1.

Nessun altro rivelatore oltre a quelli dei satelliti Konus e HEAO le ha individuate e alcuni ricercatori hanno ipotizzato che i risultati di questi satelliti siano forse dovuti a errori di calibrazione. In effetti è difficile calibrare l'efficienza di un rivelatore nella banda compresa tra 40 e 70 KeV perché la risposta è complicata dal comportamento dello iodio che esso contiene. Un'analisi approfondita degli esperimenti effettuati con i satelliti Konus e HEAO-1 indica però che la calibrazione è probabilmente corretta. Sembra quindi possibile ipotizzare che le righe spettrali appartengano davvero agli impulsi.

L'ipotesi più diffusa sull'origine delle righe è quella che le attribuisce a «emissione di ciclotrone», un tipo di radiazione che si produce a frequenze molto ben determinate quando gli elettroni si muovono rapidamente con moto circolare in un forte campo magnetico.

Secondo un'altra spiegazione, la radiazione di un impulso proviene da due sorgenti molto vicine con spettri diversi. In questa prospettiva, la separazione angolare tra le due emittenti, vista dalla Terra, sarebbe così piccola da far sembrare che la radiazione provenga da un'unica sorgente. È possibile che due spettri diversi, sommati, ne producano un terzo che presenti questo tipo di righe.

Mentre le informazioni sullo spettro si possono utilizzare per determinare il meccanismo all'origine degli impulsi, quelle relative all'intensità degli impulsi osservati possono servire a stabilire se le sorgenti siano distribuite in modo isotropo in tutto lo spazio. Uno dei metodi consiste nel tracciare un grafico del numero di impulsi che superano un qualsiasi livello di luminosità dato *(si veda l'illustrazione a pagina 35)*. Su un diagramma logaritmico i punti dovrebbero disporsi lungo una retta con una pendenza ben precisa se gli impulsi sono distribuiti uniformemente. Se, invece, molte delle sorgenti di impulsi gamma si trovano a una distanza particolare dalla Terra, il grafico presenterà un'irregolarità all'intensità corrispondente alle sorgenti osservate proprio a quella distanza.

Verso le basse intensità, in effetti, il grafico si discosta dalla retta, anche se i punti sono troppo pochi per essere sicuri che la deviazione sia statisticamente significativa. La spiegazione più facile dell'andamento osservato per la curva è che le sorgenti vicine siano distribuite uniformemente mentre quelle lontane sarebbero in qualche modo raggruppate nello spazio, lasciando grandi regioni prive di sorgenti o quasi. Una situazione di questo genere potrebbe svilupparsi naturalmente se le emittenti fossero distribuite all'interno di una regione di forma discoidale della nostra galassia. Purtroppo, però, anche molti altri tipi di distribuzione si conformerebbero ai dati osservati.

Inoltre il grafico sembra in contraddizione con ciò che sappiamo sulle posizioni degli impulsi osservati. Infatti esso sottintende che le sorgenti di impulsi gamma lontane non siano distribuite nello spazio in maniera uniforme; quindi esse dovrebbero apparire concentrate nel cielo in certe direzioni (per esempio sul piano galattico o verso un qualche ammasso di galassie vicino). Però, come ha osservato nel 1981 Mark Jennings dell'Università della California a Riverside, gli impulsi deboli sembrerebbero distribuiti in maniera uniforme. Fino a oggi per risolvere questo problema sono state proposte cinque spiegazioni, tutte accettabili ma reciprocamente incompatibili.

Fin qui il quadro che ho dato della situazione è ben poco allegro: buona parte dei dati osservativi fondamentali sono viziati da carenze degli strumenti di rilevamento, e anche ammesso che siano accettabili sono immancabilmente suscettibili di numerose spiegazioni diverse. A questo punto diventa ragionevole chiedersi se davvero possediamo qualche conoscenza non banale sulla natura delle sorgenti di impulsi gamma.

Per fortuna la situazione non è forse così drammatica come l'ho descritta. Negli ultimi anni i ricercatori sono giunti a un accordo riguardo a determinate proprietà di tali sorgenti. Esso è centrato su tre punti fondamentali: un sistema a impulsi gamma contiene una stella di neutroni; questa stella possiede un intenso campo magnetico; la maggior parte delle emittenti osservate, infine, si trova all'interno della Galassia. Anche se nessuno di questi tre punti è stato dimostrato in maniera definitiva, e ciascuno è suscettibile di contestazioni valide, prese tutte insieme le tre ipotesi forniscono la spiegazione più semplice delle indicazioni oggi disponibili.

Chiave di volta dell'accordo è la convinzione che nell'impulso di raggi gamma sia in qualche modo implicata una stella di neutroni, un corpo molto piccolo e denso formatosi nelle ultime fasi dell'evoluzione stellare. (Il raggio di una stella di neutroni è pari a una decina di chilometri, mentre la densità centrale può superare i 60 milioni circa di tonnellate per centimetro cubo.) La gravità superficiale e il campo magnetico intensissimi di questo tipo di oggetto contengono senza dubbio energia sufficiente a generare un impulso di raggi gamma; sono stati proposti molti meccanismi in grado di convertire l'una nell'altro. Un altro motivo che spiega l'interesse dei teorici per modelli che comprendono stelle di neutroni è che queste stelle esistono sicuramente e sono relativamente diffuse nella nostra galassia.

Queste considerazioni teoriche mostrano come sia plausibile che una sorgente di impulsi gamma contenga una stella di neutroni; alcuni dati osservativi sottolineano la probabilità che ciò avvenga effettivamente. Uno di questi dati è il brevissimo periodo di tempo entro cui ha luogo la variazione dell'intensità degli impulsi: alcuni sono durati solo 0,01 secondi, e uno, che è avvenuto il 5 marzo 1979, ha raggiunto la massima intensità in 0,0002 secondi. Dal momento che una sorgente non può variare di molto la propria luminosità in un tempo inferiore a quello che è necessario alla luce per attraversare la regione in cui l'emissione ha luogo, l'emittente che ha generato l'impulso del 5 marzo 1979 deve avere un'estensione inferiore agli 0,0002 secondi luce, pari a oltre 60 chilometri. Esistono pochi oggetti astronomici che soddisfino questo limite e pochi che dispongano di un'energia sufficiente a generare un impulso: una stella di neutroni risponde a entrambi questi requisiti.

Un altro elemento osservativo a favore del legame tra stelle di neutroni e impulsi gamma si basa su una particolarità individuata nel 7 per cento circa degli spettri di questi ultimi: una riga a circa 420 KeV. Esistono pochi processi capaci di emettere una riga marcata a un'energia così alta; l'ipotesi più plausi-

Le caratteristiche insolite dell'impulso del 5 marzo risultano evidenti dall'esame della sua curva di luce. L'intensità luminosa ha raggiunto in meno di 0,0002 secondi un massimo di oltre 10 volte superiore a quello di qualunque altro impulso noto. Nella fase di decadimento invece l'intensità oscillava con un periodo di circa otto secondi; è l'unico caso di periodicità riscontrato in un impulso gamma.

Gli spettri degli impulsi gamma forniscono alcuni elementi per capire la natura del fenomeno. Alcuni presentano manifestazioni di assorbimento o di emissione tra 40 e 70 KeV (*d, e, f*), forse dovute a elettroni che si muovono rapidamente di moto circolare in un forte campo magnetico vicino alla sorgente dell'impulso. Un'altra caratteristica evidente di alcuni spettri (*b, c, d*) è una riga di emissione intorno a 420 KeV, che potrebbe essere causata da collisioni tra elettroni e positoni.

bile è che essa si formi quando si ha una collisione tra gli elettroni e le loro antiparticelle, i positoni. Questo fenomeno provocherebbe l'annichilazione totale di ambedue i tipi di particelle e la conversione della loro energia di massa in due raggi gamma. Normalmente ciascun raggio gamma avrebbe un'energia di 511 KeV; se però l'annichilazione avesse luogo vicino alla superficie di una stella di neutroni, per raggiungere la Terra i raggi gamma dovrebbero prima uscire dal pozzo di energia gravitazionale della stella di neutroni, perdendo energia. (Questa diminuzione di energia è lo spostamento gravitazionale verso il rosso previsto dalla teoria della relatività generale). In realtà i raggi gamma perderebbero esattamente l'energia necessaria a colmare la differenza tra i 511 KeV ai quali erano stati emessi nella collisione tra elettrone e positone e i 420 KeV della riga di emissione osservata in alcune sorgenti di impulsi gamma.

Un'ulteriore indicazione a favore dell'idea che in queste sorgenti siano attive delle stelle di neutroni proviene da due caratteristiche uniche dell'impulso del 5 marzo 1979. Una è la notevole vicinanza tra la posizione dell'impulso nel cielo e quella di un resto di supernova. Accettando per buona la corrispondenza è probabile che la sorgente dell'impulso sia legata a un resto di supernova. Dal momento poi che spesso, nelle esplosioni delle supernove, si formano stelle di neutroni, è ragionevole supporre che responsabile dell'impulso del 5 marzo sia almeno in parte una stella di neutroni.

La seconda caratteristica unica di questo impulso è che l'intensità oscillava con un periodo di circa otto secondi. Molte emissioni astronomiche periodiche si possono spiegare con la rotazione di una stella, durante la quale un fascio di radiazione, emesso da una regione della sua superficie, colpisce la Terra periodicamente. Un periodo di rotazione di otto secondi è troppo breve per quasi tutti i tipi di stella, ma è caratteristico delle stelle di neutroni.

Il secondo punto di accordo raggiunto negli ultimi tempi sui modelli delle sorgenti di impulsi gamma è che un intenso campo magnetico potrebbe far parte del fenomeno. Ma, anche qui, nessuna delle indicazioni a favore di questa ipotesi risulta di per sé convincente; ognuna può essere interpretata senza fare ricorso ai campi magnetici. Nel complesso, però, tutte offrono una base accettabile per inserire un campo magnetico intenso nella teoria delle sorgenti di impulsi gamma.

Una delle considerazioni più convincenti è quella che si basa sull'osservazione di righe di emissione o di assorbimento nella regione compresa tra 40 e 70 KeV. Se, davvero, tali righe sono dovute alla rotazione degli elettroni, il campo magnetico che li costringe a percorrere traiettorie circolari deve essere dell'ordine di 10^{12} gauss. (Quello terrestre, per confronto, è di circa 0,5 gauss.) Se esso fosse molto più debole, gli elettroni emetterebbero radiazione di ciclotrone a energia inferiore.

Una seconda considerazione desunta dalle osservazioni si basa sulla periodicità di otto secondi, riscontrata nell'impulso del 5 marzo 1979. Questa modulazione era causata probabilmente dalla rotazione di una stella di neutroni: una regione emittente della stella, ruotando, entrava e usciva periodicamente dal campo visivo della Terra. Questa regione non poteva trovarsi a uno dei poli della stella, dove l'asse di rotazione attraversa la superficie, perché altrimenti non avrebbe ruotato e quindi non si sarebbe mossa attraverso la visuale della Terra. Deve quindi esistere un meccanismo, non simmetrico rispetto all'asse di rotazione, che determini la posizione della regione emittente sulla superficie della stella. Un probabile responsabile di questa «rottura di simmetria» è un campo magnetico con l'asse non allineato rispetto a quello di rotazione della stella di neutroni.

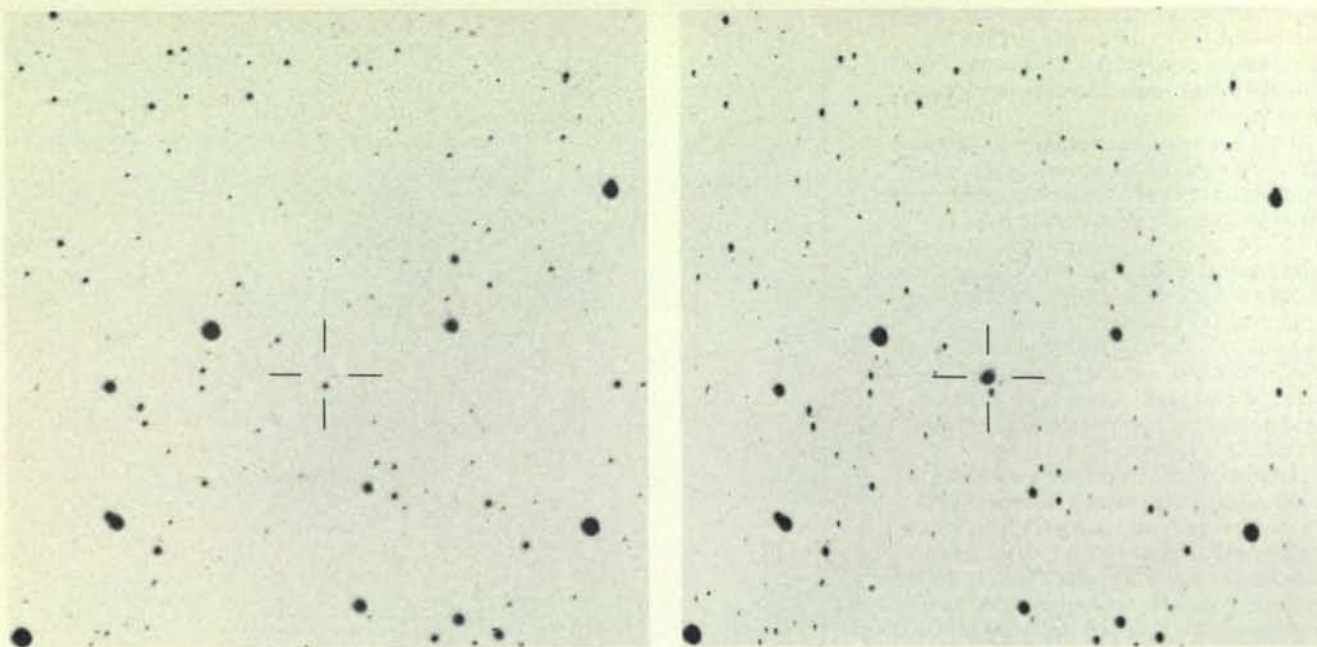
L'ipotesi della presenza di un campo magnetico in una sorgente di impulsi gamma è sostenuta da considerazioni teoriche oltre che dedotte da osservazioni. Una delle più valide tra queste è quella secondo cui qualsiasi regione, sede di un'intensa emissione gamma, deve essere confinata da una forza di qualche genere, altrimenti la pressione di radiazione prodotta dai raggi gamma la farebbe espandere molto rapidamente. Neppure la formidabile gravità superficiale di una stella di neutroni è in grado di confinare una sorgente gamma intensa, mentre un campo magnetico intorno ai 10^{12} gauss sarebbe probabilmente sufficiente.

Il terzo punto su cui ultimamente si è raggiunto un accordo è l'idea che quasi tutte le sorgenti di impulsi gamma si trovino nella nostra galassia. Una delle principali considerazioni a favore di questa ipotesi si basa sulla grande quantità di energia che sarebbe necessario attribuire a tali sorgenti se fossero extragalattiche. Per raggiungere la luminosità osservata sulla Terra un'emittente galattica dovrebbe produrre circa 10^{38} erg di energia, mentre una sorgente extragalattica dovrebbe produrne almeno 10^{46} per ottenere lo stesso effetto. Mentre però è relativamente facile elaborare modelli di processi in grado di produrre 10^{38} erg, la realizzazione di un modello capace di generare addirittura 10^{46} erg comporta molte gravi difficoltà.

Una considerazione tratta dalle osservazioni e basata sul grafico del numero di impulsi che superano una data intensità appoggia anche l'ipotesi che le loro sorgenti si trovino nella Galassia. Se, infatti, le sorgenti osservate fossero per lo più extragalattiche, dovrebbero molto probabilmente trovarsi all'interno di altre grandi concentrazioni di massa (per esempio in altre galassie o ammassi di galassie). La loro distribuzione spaziale, così come è stata osservata, sarebbe perciò irregolare. Esisterebbero grandi gruppi di sorgenti riuniti a distanze diverse dalla Terra e, in tal caso, il grafico presenterebbe una serie di irregolarità in corrispondenza delle distanze dei vari aggregati, irregolarità che, invece, non compaiono.

La distribuzione delle sorgenti di impulsi gamma nello spazio si può valutare tracciando un grafico del numero di impulsi più intensi di una certa intensità limite in funzione di questa. Se la distribuzione fosse uniforme, il grafico, disegnato su scala logaritmica, dovrebbe essere una linea retta con pendenza -3/2 *(in colore)*. La deviazione della curva effettiva dalla retta indica che le emittenti osservate forse non sono distribuite in maniera uniforme. Le linee tratteggiate rappresentano cataloghi di dati compilati sulla base di risultati di singoli satelliti.

L'accordo attuale è un buon punto di partenza, ma la sua utilità teorica è limitata dalle difficoltà insite nel tentativo di ottenere dati validi da impulsi di fotoni ad alta energia quali sono appunto i raggi gamma. Gli astrofisici si sono resi conto ormai da molto tempo che sarebbe possibile superare molte di queste difficoltà se esistesse un sistema per osservare fotoni provenienti dalle sorgenti di impulsi gamma, ma di energia inferiore. A tal fine sono stati compiuti molti sforzi nel determinare le posizioni precise del maggior numero possibile di impulsi. Queste posizioni particolari sono poi state esaminate con telescopi ottici per vedere se qualche tipo di fenomeno ottico fosse legato agli impulsi stessi.

I risultati di questa ricerca sono stati immancabilmente negativi, con due eccezioni. La prima è stata l'impulso insolito del 5 marzo 1979, identificato provvisoriamente con i resti di una superno-

In una delle due fotografie realizzate il 17 novembre 1928 si vede un lampo ottico; la fotografia a sinistra è stata ottenuta 45 minuti prima dell'altra. Il lampo è avvenuto nella stessa regione del cielo dalla quale è provenuto un impulso gamma il 19 novembre 1979. Le immagini delle stelle nella fotografia a destra sono allungate perché il telescopio non seguiva esattamente il moto del cielo. Il lampo, invece, non presenta una scia, e questa caratteristica sta a indicare che la sua durata è stata nettamente inferiore al tempo di esposizione totale.

va. La seconda è costituita da tre lampi ottici da me osservati nel 1981 mentre frequentavo il Massachusetts Institute of Technology (MIT). Nel tentativo di individuare lampi ottici provenienti da sorgenti di impulsi gamma ho esaminato circa 30 000 fotografie, che costituiscono una parte dei 500 000 esemplari dell'archivio dello Harvard College Observatory, e ho analizzato tutte quelle che riproducevano una zona del cielo in cui era stato osservato in seguito un impulso gamma. In tre fotografie (di altrettante regioni diverse) che risalgono rispettivamente al 1901, al 1928 e al 1944 ho individuato immagini di tipo stellare non visibili in altre fotografie delle stesse zone. Queste immagini, prodotte da lampi ottici, sono sicuramente legate agli impulsi gamma osservati in seguito nelle stesse posizioni.

Anche se sarebbe utile disporre di ulteriori dati, già da questi tre lampi si possono ricavare molte informazioni. In tutti e tre i casi, per esempio, l'energia dell'impulso gamma recente è quasi esattamente 1000 volte superiore a quella dell'impulso ottico. Insieme con le varie considerazioni teoriche, questo valore può servire a mostrare come la stella di neutroni che causa gli impulsi debba avere una compagna di qualche genere, forse una stella debole o un disco di materia più fredda.

Dal momento che le posizioni delle sorgenti di impulsi si possono determinare con maggiore precisione partendo da dati nel visibile più che da quelli relativi alla radiazione gamma, le tre posizioni stabilite nel visibile ci hanno dato modo di condurre ricerche approfondite di sorgenti di impulsi quiescenti mediante grandi telescopi ottici. I risultati non permettono di raggiungere conclusioni definitive per le tre posizioni, ma sembra che le sorgenti debbano essere molto deboli quando non emettono impulsi.

L'analisi statistica dei tre lampi ottici rivela che le sorgenti di impulsi gamma debbono emettere tali lampi in media una volta all'anno: ho esaminato foto d'archivio per una media di tre mesi di lavoro per ognuna delle 12 posizioni nelle quali si sa che esistono sorgenti di impulsi gamma; in pratica è come se avessi analizzato foto per un'esposizione cumulativa di tre anni. In questo complesso di fotografie ho individuato tre lampi, il che indica che l'intervallo medio tra i lampi ottici di una sorgente di questo tipo dev'essere di circa un anno. Questo risultato servirà forse a scartare quei modelli che non permettono il verificarsi di tali lampi con cadenza annuale.

La conseguenza forse più importante dell'osservazione dei tre lampi è però il fatto che essi costituiscono un potente monito, ricordando agli astrofisici che l'osservazione delle sorgenti di impulsi gamma a energie inferiori è non solo possibile, ma anche opportuna.

Nei prossimi anni si assisterà all'avvio di diversi esperimenti potenzialmente molto utili per raccogliere nuove informazioni sulle emittenti di impulsi gamma. Il più entusiasmante è quello con il rivelatore di impulsi a bordo del satellite Gamma Ray Observatory, il cui lancio è previsto per il 1988. Questo strumento, senza essere essenzialmente diverso dai suoi predecessori tranne per la sensibilità, molte volte superiore, rivelerà impulsi da 10 a 100 volte più deboli di quelli accessibili all'esperimento Konus, individuandone inoltre la posizione con l'approssimazione di un grado. Una particolare modifica del rivelatore registrerà inoltre spettri su una vasta banda di energie per ciascun impulso osservato.

Un altro importante esperimento, volto alla ricerca di lampi ottici e oggi vicino alla conclusione della fase di allestimento, sarà costituito da due parti. La prima, curata da Ronald Vanderspek e George R. Ricker, Jr., del MIT, prevede l'osservazione di quasi tutto il cielo nel visibile in attesa di aumenti improvvisi di luminosità. La posizione di qualsiasi lampo verrà comunicata entro un secondo dal suo inizio alla seconda componente dell'esperimento, un telescopio ottico in grado di puntare qualsiasi regione del cielo entro un secondo. (Questa componente è in corso di realizzazione da parte di Bonnard J. Teegarden, Ravi Kaipa, Tycho T. von Rosenvinge e Cline presso il Goddard Space Flight Center.) I ricercatori prevedono di rivelare circa una ventina di impulsi ottici all'anno.

I risultati che ci aspettiamo da questi e da altri esperimenti dovrebbero chiarire finalmente la causa all'origine degli impulsi gamma. Nonostante l'accordo attuale, infatti, questi impulsi restano una delle classi di fenomeni astrofisici più misteriose: non vedo l'ora che siano disponibili i dati che ci permetteranno di comprenderli.

Gli acciai di Damasco

Un tenore di carbonio più alto rispetto agli acciai attuali e una fucinatura a regola d'arte erano le caratteristiche delle lame di Damasco, famose per la resistenza e la tenacità oltre che per la bellezza delle venature

di Oleg D. Sherby e Jeffrey Wadsworth

Nel suo romanzo storico sulle avventure dei crociati in Palestina, *Il Talismano*, Sir Walter Scott narra di un incontro tra Riccardo Cuor di Leone e il re saraceno Saladino. I due contendenti stanno vantando i pregi delle rispettive spade. Per dimostrare la potenza della sua arma, Riccardo afferra a due mani la pesante spada a lama dritta e fende di colpo una mazza di acciaio. È la volta del Saladino che, afferrato un cuscino di seta, lo avvolge attorno alla sua scimitarra e «sguaina l'arma attraverso il cuscino...con una tal leggerezza che questo sembra sfaldarsi naturalmente, senza traccia di violenza». I crociati non credendo ai loro occhi pensano che si tratti di un trucco, ma il Saladino li sbalordisce di nuovo dividendo in strisce con la scimitarra un velo che volteggia leggero nell'aria. Stando alla descrizione di Walter Scott, l'arma del Saladino, meravigliosamente affilata e leggera, aveva «una lama stretta e ricurva che, diversamente da quelle dei franchi, non scintillava, aveva un colore blu cupo e recava un'infinità di venature».

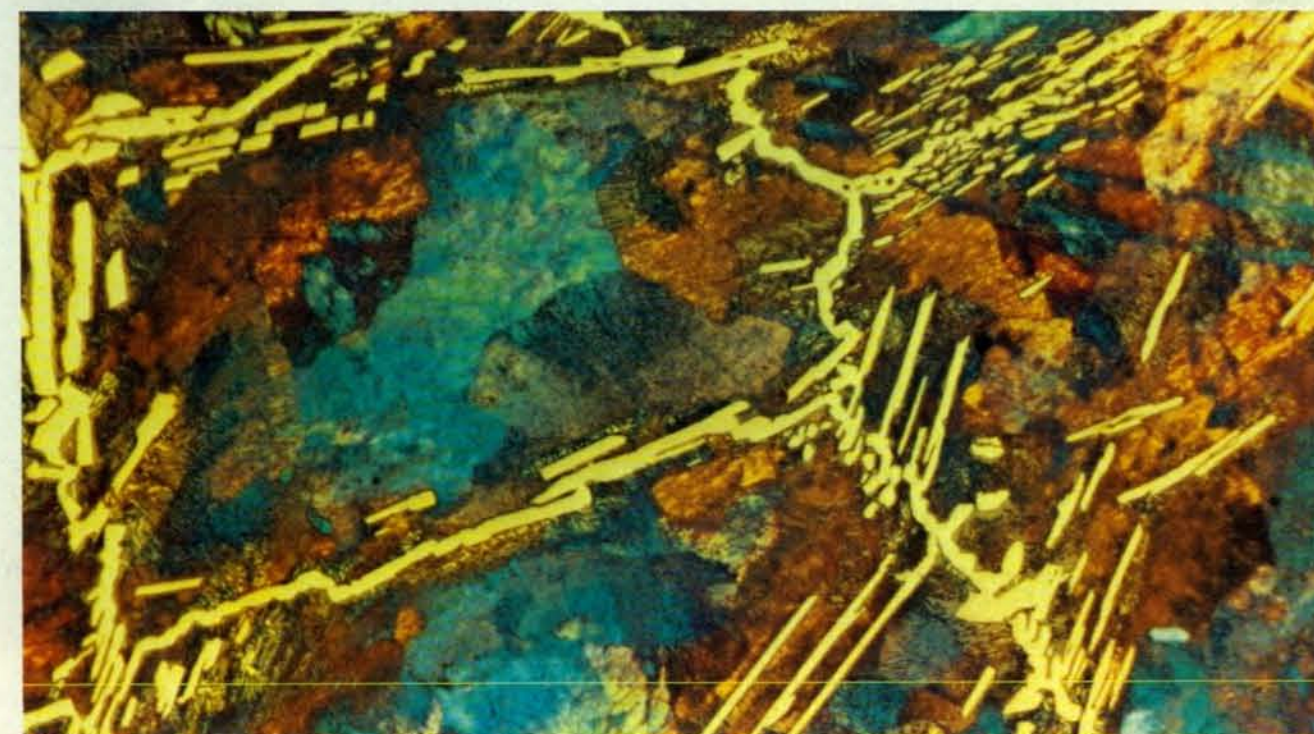
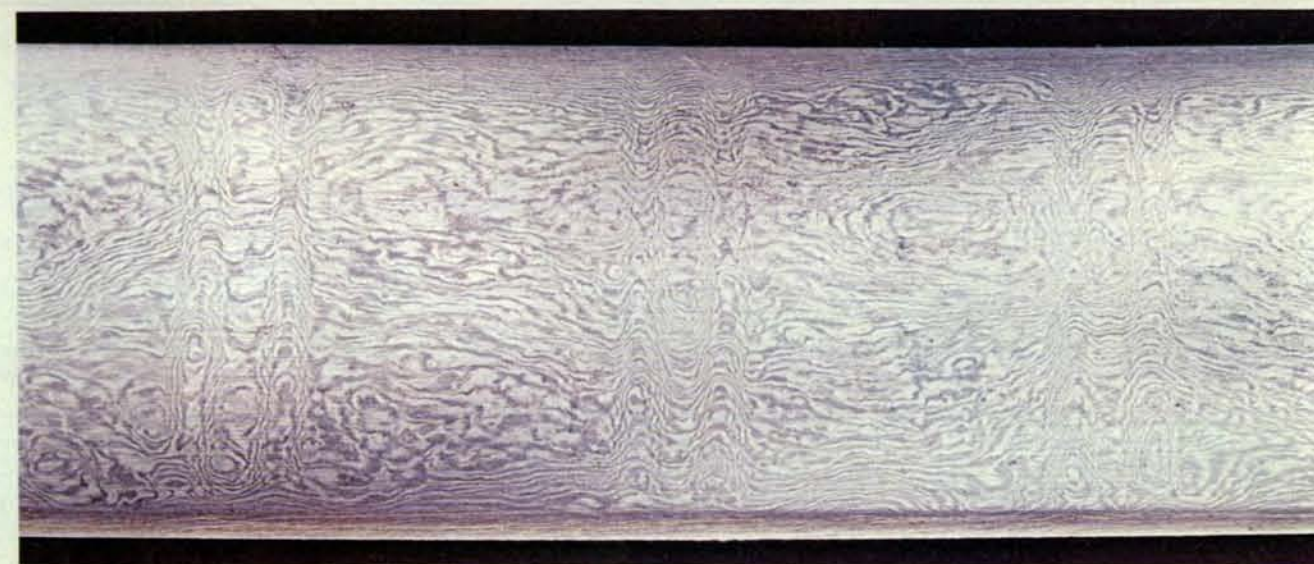
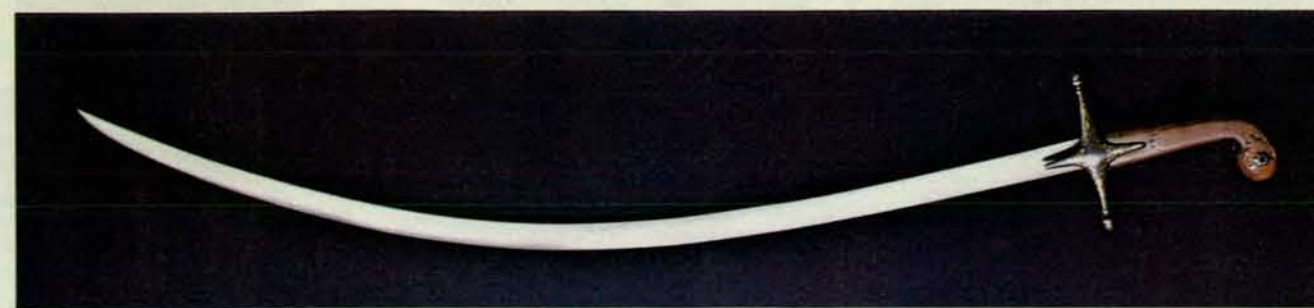
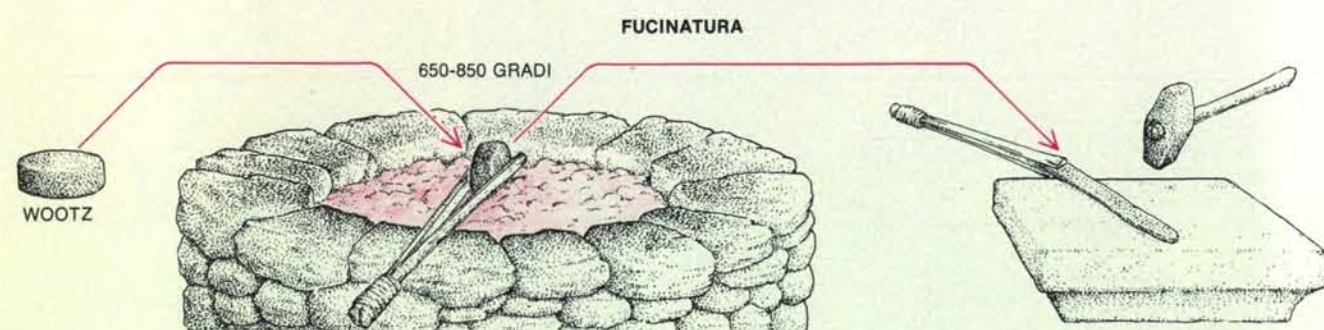
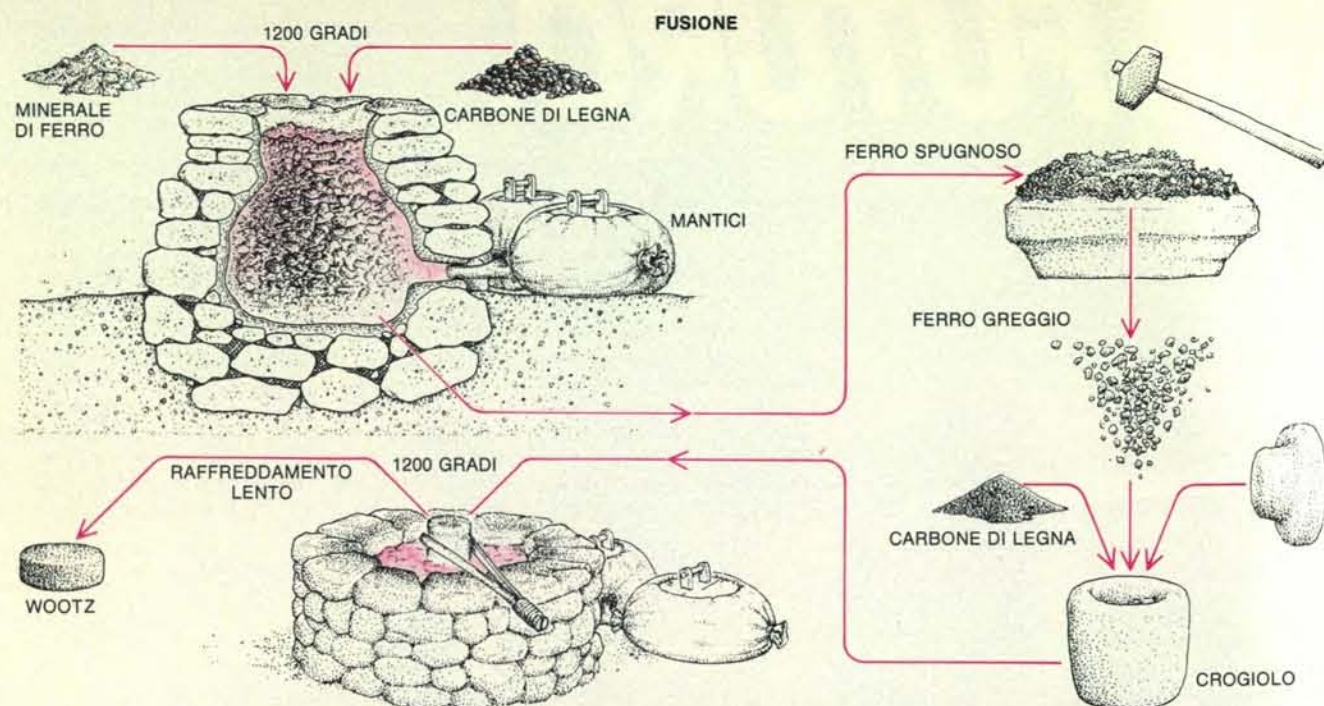
Sebbene questa descrizione non sia priva di qualche licenza poetica (per esempio non era possibile che il Saladino avesse una scimitarra in quanto le lame ricurve fecero la loro comparsa solo diversi secoli dopo il 1192, anno in cui si sarebbe svolto l'incontro con Riccardo), fornisce però un'immagine sostanzialmente fedele del tipo di lama usato nell'Islam ai tempi del Saladino. Erano lame eccezionalmente dure all'impatto - in grado cioè di conservare in ogni caso l'affilatura - e a un tempo tenaci quanto bastava per assorbire i colpi durante il combattimento senza rompersi. Quelle lame dovevano le loro qualità e gli splendidi motivi delle venature superficiali al materiale di cui erano fatte: l'acciaio di Damasco. Fin dal tempo delle Crociate le spade e le armature di Damasco sono state materia di leggenda. Per secoli, nelle epoche successive, sono state oggetto di fascino e di disperazione per i fabbri europei che invano tentavano di riprodurne fedelmente la damascatura.

La scoperta del segreto delle spade di Damasco divenne una sfida che suscitò anche l'attenzione di eminenti scienziati europei. Uno di questi fu Michael Faraday, figlio di un fabbro, che dopo aver analizzato nel 1819 (prima di inventare il motore e il generatore elettrici) un acciaio di Damasco attribuì le sue straordinarie caratteristiche alla presenza di piccole quantità di silice e allumina. Benché le conclusioni fossero sbagliate, la relazione di Faraday stimolò Jean Robert Bréant, ispettore alla zecca di Parigi, a intraprendere una serie di esperimenti per valutare l'effetto dell'aggiunta di vari elementi all'acciaio. Fu appunto Bréant che individuò per primo, nel 1821, quale fosse, dal punto di vista metallurgico, la caratteristica peculiare degli acciai di Damasco, ossia la straordinaria durezza, la tenacità e la bellezza dovute a un elevato tenore di carbonio. Bréant identificò come «acciaio carburato» il costituente delle zone pallide della superficie damascata e definì semplicemente «acciaio» la matrice scura.

Bréant riuscì a fabbricare spade damascate, ma non rivelò mai nei particolari il procedimento da lui adottato, del quale forse gli sfuggiva anche l'importanza di alcune fasi. La base per una piena comprensione scientifica dell'acciaio di Damasco non si raggiunse che al volgere del secolo, quando un buon numero di ricercatori riuscì a svelare le transizioni di fase che gli acciai subiscono in funzione della temperatura e del tenore di carbonio. Ancor oggi, pur essendo il diagramma ferro-carbonio ormai interamente noto, l'arte di fabbricare spade di Damasco rimane, per la legge americana, coperta da brevetto.

L'interesse che coltiviamo per questo argomento deriva dalle nostre ricerche sui moderni acciai ad altissimo carbonio. Questi acciai, che hanno un tenore di carbonio compreso tra l'1 e il 2,1 per cento, hanno scarsissima applicazione commerciale perché sono ritenuti fragili. Le spade di Damasco, però, pur avendo un tenore di carbonio variante dall'1,5 al 2 per cento, godevano di una indiscussa reputazione di tenacità; questo suggerisce che la fragilità solitamente indotta da un'elevata percentuale di carbonio può essere evitata con trattamenti adeguati. Alla Stanford University siamo riusciti a ottenere acciai ad altissimo carbonio che al pari delle spade di Damasco non solo presentano a temperatura ordinaria un'elevata resistenza a rottura, ma sono anche duttili. Siamo pure riusciti a riprodurre il leggendario acciaio damascato. I procedimenti da noi seguiti in laboratorio sono essenzialmente gli stessi adottati per la prima volta nelle antiche fucine del Vicino Oriente.

La più antica descrizione delle lame di Damasco risale al 540 d.C., ma non è da escluderne l'impiego fin dai tempi di Alessandro Magno (circa 323 a.C.). Il loro nome non deriva dal luogo di origine, ma dal fatto che a Damasco furono viste per la prima volta dai crociati. L'acciaio di cui erano fatte proveniva in realtà dall'India dove, con il nome di *wootz* (acciaio fuso in crogiolo), era oggetto di commercio su vasta scala, sotto forma di pani di dimensioni analoghe a quelle di un disco da hockey. Si ritiene che le migliori lame venissero ottenute in Persia fucinando del wootz proveniente dall'India con il quale si fabbricavano anche scudi e corazze. Benché la diffusione degli acciai di Damasco rispecchi geograficamente l'espansione islamica, nel Medioevo erano noti anche in Russia con il nome di *bulat*.

Come accade per tutti i procedimenti di fabbricazione dell'acciaio, per ottenere il wootz era prima necessario rimuovere l'ossigeno dal minerale di ferro, costituito da ossido; la successiva aggiunta di carbonio al ferro così ridotto ne provocava l'indurimento trasformandolo in acciaio. Il carbonio necessario veniva ricavato da carbone di legna, legna o fo-

Le venature superficiali sulla lama di una scimitarra persiana (*in alto e al centro*) rispecchiano variazioni del tenore di carbonio nell'acciaio ad altissimo carbonio: le zone chiare sono costituite da carburo di ferro (cementite) mentre il fondo scuro è costituito da ferro assai più povero di carbonio. Le venature diventavano visibili soltanto quando la lama, finita in ogni sua parte, veniva lucidata e attaccata con un acido che agiva di preferenza sulla matrice ferrosa lasciando pressoché inalterata la cementite. Il reticolo della cementite è chiaramente distinguibile nella micrografia di un acciaio ad altissimo carbonio (*in basso*); ferro e cementite, a strati alterni, compaiono anche nella struttura di fondo. Gli acciai di Damasco dovevano la loro tenacità alla fucinatura che, disperdendo la cementite, conferiva alle venature superficiali il loro aspetto definitivo. Gli effetti della fucinatura sono tuttora riconoscibili: le venature verticali della scimitarra persiana, al cui insolito andamento veniva dato il nome di «Scala di Maometto», erano il risultato di martellature localizzate. La scimitarra, risalente al XVII secolo o a epoca più tarda, è conservata al Metropolitan Museum of Art di New York. L'ingrandimento della micrografia è di circa 200 volte.

Il processo di fabbricazione di una spada di Damasco iniziava con la fusione di un acciaio ad altissimo carbonio, detto *wootz*, nelle fonderie dell'India. Minerale di ferro e carbone di legna venivano mescolati e riscaldati a 1200 gradi centigradi in bassoforni di pietra dove il minerale di ferro, reagendo con il carbonio proveniente dal carbone di legna, veniva ridotto (ossia privato di ossigeno) e assumeva una consistenza spugnosa. Con colpi ripetuti di martello le impurezze venivano espulse dalla massa spugnosa e si otteneva ferro greggio in pezzatura minuta e con basso tenore di carbonio. Il contenuto di carbonio veniva poi aumentato riscaldando ad alta temperatura ferro greggio e carbone di legna in crogioli a chiusura ermetica per impedire la riossidazione del ferro da parte dell'aria. Quando scuotendo il crogiolo si udiva uno sguazzo che indicava la presenza di materiale fuso, si lasciava raffreddare lentamente il crogiolo mantenendolo nel forno. Il wootz veniva posto in commercio sotto forma di pani di parecchi centimetri di diametro. Nelle fucine del Vicino Oriente si fabbricava una lama da ciascun pane, riscaldandolo a 650-850 gradi, ossia l'intervallo di temperature in cui gli acciai ad altissimo carbonio sono duttili. Le lame così ottenute venivano temprate mediante riscaldamento e brusco raffreddamento in acqua, salamoia o altri liquidi.

gliame. Tipicamente il procedimento consisteva nel mescolare minerale di ferro e carbone di legna e portarli a una temperatura di circa 1200 gradi centigradi in un forno di pietra. L'ossigeno veniva separato dal minerale facendolo reagire con il carbonio del carbone di legna. Dalla quantità di carbone presente nella miscela dipendeva se il risultato dell'operazione era ferro greggio a tenore di carbonio molto basso oppure ghisa con tenore di carbonio superiore al 4 per cento. Per ottenere alla fine il wootz gli artigiani indiani procedevano in due modi: arricchivano in carbonio il ferro greggio oppure decarburavano la ghisa.

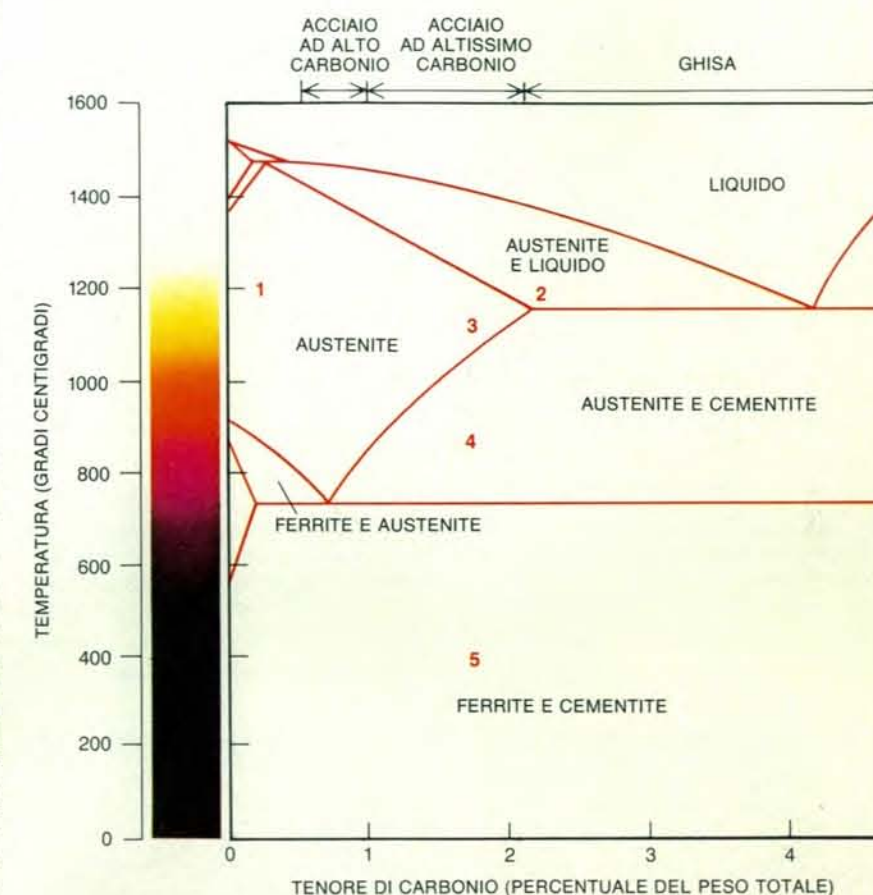
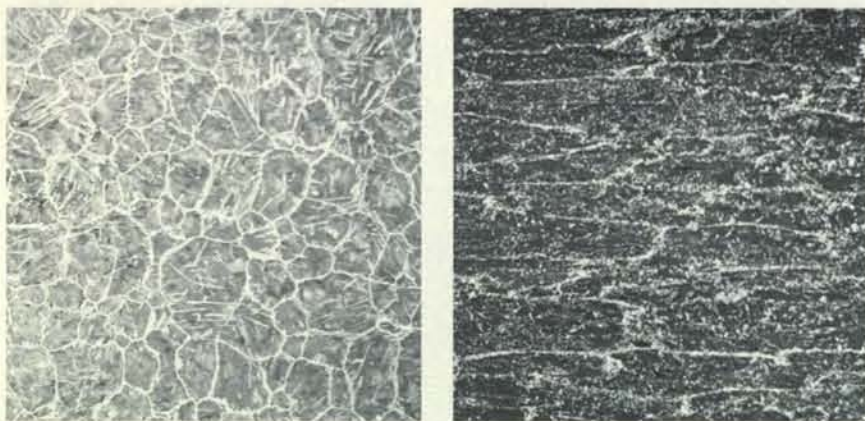
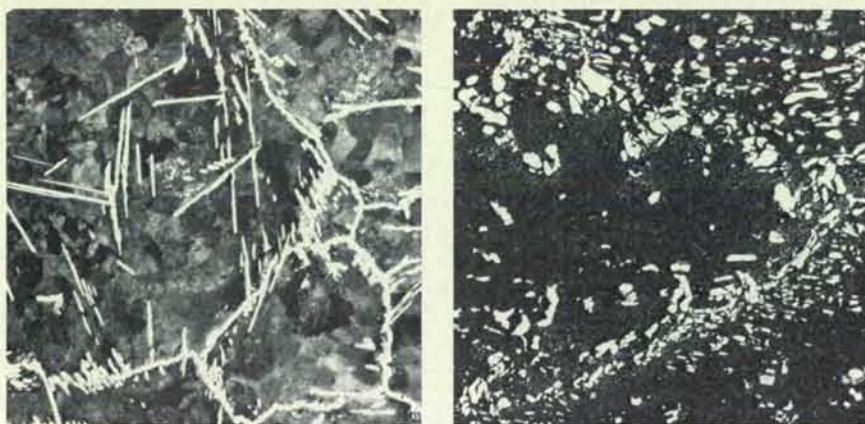
Il procedimento seguito per ottenere wootz da ferro greggio è meglio conosciuto di quello che partiva da ghisa. Il ferro in pezzatura minuta veniva mescolato con carbone di legna in un crogiolo d'argilla a chiusura ermetica, di circa otto centimetri di diametro e 15 di altezza, che poi veniva riscaldato a circa 1200 gradi. A questa temperatura la fusione del ferro non è ancora iniziata, ma la sua struttura cristallina, cubica a facce centrate, consente agli atomi di carbonio di trovar posto tra gli atomi di ferro. Aveva luogo così una graduale diffusione di carbonio nel ferro e si formava la lega che oggi chiamiamo austenite.

L'aggiunta di carbonio abbassava il punto di fusione del metallo e non appena la superficie dei pezzi di ferro greggio, arricchendosi in carbonio, arrivava a contenerne il 2 per cento circa, un sottile strato bianco di ghisa liquida cominciava a formarsi sulla superficie stessa. Quando, scuotendo il crogiolo, si avvertiva lo sguazzo provocato dalla presenza di materiale fuso, si aveva l'indicazione che il carbonio si era ormai disciolto nel ferro in quantità apprezzabili.

A quel punto il crogiolo veniva lasciato raffreddare molto lentamente, talvolta per alcuni giorni. Ciò permetteva al carbonio di distribuirsi uniformemente nell'acciaio, in quantità compresa tra l'1,5 e il 2 per cento. Allorché la temperatura della massa metallica scendeva sotto i 1000 gradi circa, una parte del carbonio precipitava dalla soluzione e si formava un reticolo di carburo di ferro o cementite (Fe_3C) attorno ai grani di austenite. La lentezza del raffreddamento causava un sensibile accrescimento del grano austenitico e così il reticolo della cementite era alquanto grossolano.

Appunto alla presenza di quel reticolo grossolano era dovuto il caratteristico andamento delle venature visibili sulle lame di Damasco. La cementite ha, però, anche caratteristiche negative: pur essendo molto dura, a temperatura ambiente è anche estremamente fragile. La naturale fragilità sarebbe poi risultata ancora più dannosa per la presenza della struttura a reticolo, che predisponeva un comodo tracciato per la propagazione di eventuali cricche. In realtà le spade di Damasco erano tutt'altro che fragili e avevano un'eccellente tenacità che era raggiunta nel corso della fucinatura allorché il reticolo della cementite veniva frantumato con ripetute martellature.

A quanto sembra la fucinatura degli acciai di Damasco per fabbricare lame avveniva a temperatura relativamente bassa. Nel Medioevo non era possibile una misura precisa della temperatura di una fornace o di una forgia e la si valutava in base al colore del metallo che veniva lavorato. Gli acciai venivano solitamente fucinati nella fascia di temperature che scendeva dal calor bianco (1200 gradi centigradi) fino all'arancione (900 gradi); sembra però che la fucinatura del wootz avvenisse a temperature inferiori, comprese tra il rosso ciliegia (850 gradi) e il rosso sangue (650 gradi). Temperature più elevate avrebbero provocato la ridissoluzione della cementite nell'austenite. Fucinando i pani di wootz a temperatura inferiore a 850 gradi, il reticolo della cementite veniva invece frantumato in particelle sferoidali. La cementite continuava così ad assolvere la funzione di indurente dell'acciaio, ma poiché la continuità del bordo dei grani era stata interrotta, ogni pericolo di fragilità veniva a cessare.

Le spade di Damasco testimoniano una notevole abilità nella fucinatura; nella fabbricazione delle lame, lo spessore iniziale del massello veniva ridotto da tre a otto volte. Siamo riusciti a dimostrare sperimentalmente che gli acciai ad altissimo carbonio sono effettivamente duttili e facilmente fucinabili a 850 gradi. Abbiamo sottoposto dei masselli di acciaio a diverso tenore di carbo-

Il diagramma ferro-carbonio è lo strumento fondamentale per comprendere sia le caratteristiche sia il processo di lavorazione delle lame di Damasco. Quando ferro greggio e carbone di legna venivano riscaldati, in un crogiolo, a 1200 gradi il ferro si trasforma in austenite a reticolo cubico a facce centrate (*1*). Il carbonio proveniente dal carbone poteva quindi entrare in soluzione nel ferro, abbassandone il punto di fusione. Ghisa liquida si formava sulle particelle di ferro quando il tenore di carbonio del loro strato superficiale superava il 2 per cento (*2*). Il lento raffreddamento consentiva la diffusione del carbonio nell'intera massa metallica che si trasformava così in un acciaio con un tenore medio di carbonio compreso tra l'1,5 e il 2 per cento (*3*) e nello stesso tempo favoriva una crescita grossolana del grano austenitico. Quando la temperatura della massa scendeva sotto i 1000 gradi, il carbonio precipitava al bordo dei grani come cementite (*4*). Le venature chiare erano dovute alla presenza di un reticolo grossolano di cementite. Non appena la temperatura scendeva sotto ai 727 gradi, l'austenite a facce centrate dava luogo a strati alterni di cementite e di ferrite a corpo centrato, povera di carbonio (*5*). Le lame venivano temprate riscaldandole di nuovo oltre i 727 gradi e quindi raffreddandole bruscamente: l'austenite veniva così trasformata in martensite. Nel Medioevo la temperatura del ferro e degli acciai veniva valutata dal colore che assumevano.

Gli effetti della fucinatura sulle lame di Damasco sono stati riprodotti laminando un acciaio ad altissimo carbonio. Nelle micrografie in alto la struttura dell'acciaio è ingrandita 130 volte, in quelle in basso 6,5 volte. Prima della laminazione (*a sinistra*) il reticolo chiaro della cementite è continuo e le sue dimensioni sono press'a poco uniformi in ogni direzione. Dopo la laminazione (*a destra*) il reticolo si presenta allungato nel senso della laminazione stessa e frantumato in particelle sferoidali. Di conseguenza l'acciaio risulta più tenace (meno fragile).

nio - 1,3, 1,6 e 1,9 per cento - a pesanti deformazioni riducendo di ben tre volte il loro spessore in un colpo solo, senza traccia alcuna del più piccolo indizio di criccatura. Ripetendo l'operazione su un massello di ghisa, tendenzialmente più fragile a causa del maggior tenore di carbonio (2,3 per cento), si poteva notare la formazione di criccature in corrispondenza degli spigoli.

Uno dei motivi per i quali i fabbri europei incontravano tante difficoltà quando cercavano di riprodurre le lame di Damasco, persino adoperando del wootz quale materia prima, era probabilmente l'abitudine a lavorare acciai a basso carbonio, che hanno un punto di fusione più elevato. Si può quindi pensare che tentassero di fucinare al calor bianco l'acciaio proveniente dall'India, cioè in condizioni di parziale fusione. Il risultato dei loro sforzi fu probabilmente simile a quello descritto da Bréant quando osservò come «al calor bianco [gli acciai di Damasco] si sgretolassero sotto i colpi del martello».

Dopo la fucinatura le lame di Damasco venivano di solito indurite mediante trattamento termico. Si può indurire termicamente l'acciaio portandolo oltre i 727 gradi centigradi (temperatura alla quale la ferrite a corpo centrato inizia a trasformarsi in austenite a facce centrate) e quindi immergendolo bruscamente in acqua o in un altro mezzo di raffreddamento rapido. Quando la fase austenitica degli acciai ad altissimo carbonio è lasciata raffreddare lentamente, come accadeva ad esempio nei crogioli dai quali si otteneva il wootz, si forma una struttura chiamata perlite costituita da ferrite duttile, povera di carbonio, e da cementite, che invece ne è ricca, disposte a strati alterni. Se però l'acciaio viene bruscamente raffreddato, la trasformazione dell'austenite in perlite viene bloccata. I cristalli di ferro assumono ugualmente la configurazione a corpo centrato, ma, per effetto delle tensioni che si sviluppano, sono sottoposti a uno stiramento che deforma il loro reticolo da cubico a tetragonale. Questa struttura, che è detta martensite, ha ancora la possibilità di incorporare atomi di carbonio e risulta, così, dura.

Sembra che i procedimenti seguiti dagli artigiani di epoca medioevale per il trattamento termico delle lame di Damasco fossero molteplici e che spesso dessero peso a elementi che oggi a un tecnico apparirebbero del tutto privi di consistenza. Alcuni di loro, per esempio, asserivano che le spade dovessero venir bruscamente raffreddate nell'urina di un fanciullo dai capelli rossi o in quella di «una capra di tre anni nutrita per tre giorni solo di felci». Una tra le più dettagliate descrizioni di procedimenti di indurimento dell'acciaio di Damasco (*bulat*) è stata trovata nel tempio di Balgala in Asia Minore. «Si scaldi il bulat fino a farlo apparire come il Sole che sorge nel deserto, lo si raffreddi poi fin quando il suo colore non sia quello della porpora regale e lo si lasci infine cadere nella carne di uno schiavo muscoloso...la forza dello schiavo entrerà nella spada, da lui trarrà il metallo la sua resistenza.»

Queste istruzioni possono venir interpretate anche in modo diverso. La lama doveva essere portata ad alta temperatura, forse sopra i 1000 gradi («il Sole

L'analogia tra le microstrutture della sezione trasversale di una lama di Damasco (*a sinistra*) e di un laminato di acciaio ad altissimo carbonio (*a destra*) induce a pensare che siano stati ottenuti con procedimenti analoghi. La lavorazione agisce meccanicamente sul reticolo della cementite comprimendolo; lo spazio tra gli strati è di circa 100 micrometri. Gli acciai laminati presentano una damascatura meno intricata di quella delle lame ottenute per fucinatura.

che sorge nel deserto»), e quindi raffreddata all'aria fino a circa 800 gradi («porpora regale»); da ultimo doveva essere bruscamente immersa in qualcosa di tiepido (37 gradi) e denso.

Difficilmente questo procedimento di lavorazione avrebbe dato buoni risultati. Per effetto del riscaldamento sopra i 1000 gradi la cementite si sarebbe ridisciolta nei cristalli di austenite a facce centrate e il raffreddamento lento a 800 gradi avrebbe consentito il riformarsi di un reticolo di cementite grossolano al bordo dei grani; la temperatura elevata avrebbe inoltre favorito la crescita del grano cristallino oltre i limiti accettabili. Entrambi gli effetti avrebbero ridotto la tenacità della lama. Una spada fabbricata secondo i precetti del tempio di Balgala avrebbe certo raggiunto la durezza desiderata, ma sarebbe probabilmente risultata troppo fragile per reggere l'impatto di una lama che fosse stata riscaldata solo di poco sopra ai 727 gradi prima di venir bruscamente raffreddata. Una lama siffatta sarebbe risultata a un tempo dura e tenace.

Secondo la moderna teoria metallurgica, gli acciai più resistenti e tenaci sono generalmente quelli che hanno struttura cristallina più fine. Questo porta, per ironia, a pensare che le spade migliori di Damasco fossero quelle le cui lame erano del tutto prive di damascatura. Per gli artigiani del Medioevo la presenza di damascatura rappresentava certamente un criterio di controllo di qualità: costituiva infatti la prova sia di un contenuto di carbonio elevato, che garantiva la robustezza della lama, sia di una corretta fucinatura che l'aveva resa tenace. La damascatura risultava però evidente solo a patto che le particelle di cementite fossero state grossolane e non uniformemente distribuite nell'acciaio. Lame la cui microstruttura fosse stata così fine da non provocare alcun visibile contrasto in superficie si sarebbero tuttavia rivelate più resistenti e tenaci.

Per verificare la validità delle nostre teorie sulla composizione e sulla fabbricazione degli acciai di Damasco, abbiamo tentato di riprodurre in laboratorio la damascatura. In primo luogo abbiamo mantenuto un piccolo getto di acciaio, con un tenore di carbonio dell'1,7 per cento, a 1150 gradi (giallo chiaro) per 15 ore. Il riscaldamento prolungato ha favorito la dissoluzione del carbonio e la formazione di una fase austenitica a cristalli molto grossolani. Abbiamo quindi raffreddato il getto a una velocità di circa 10 gradi all'ora. Il lento raffreddamento ha fatto sì che al bordo dei grani di austenite si formasse un reticolo ininterrotto di cementite.

Dopo aver nuovamente riscaldato il getto a 800 gradi, lo abbiamo laminato riducendone di otto volte lo spessore. Questa operazione, che doveva simulare la fucinatura, ha provocato uno stiramento dei grani nella direzione di laminazione e ha frantumato il reticolo di cementite. Attaccando l'acciaio con un acido, che agiva di preferenza sulla matrice ferrosa lasciando pressoché la cementite inalterata, la damascatura era visibile a occhio nudo. La microstruttura del getto risultava sorprendentemente simile a quella degli acciai di Damasco (*si veda l'illustrazione in basso nella pagina a fronte*).

Il procedimento descritto rappresenta solo uno dei possibili metodi di fabbricazione dell'acciaio di Damasco. Ne esistevano probabilmente molti altri. Non è escluso che nelle fucine del Vicino Oriente siano stati prodotti persino acciai ad altissimo carbonio privi di damascatura. Proprio questo siamo riusciti a ottenere nel nostro laboratorio laminando acciaio fuso non appena attraversata la fase austenite più cementite, durante il suo raffreddamento dai 1100 gradi iniziali. Poiché l'azione meccanica aveva affinato il grano austenitico e provocato la precipitazione della cementite in particelle fini e uniformemente distribuite, risultò impedita la formazione di qualsiasi contorno grossolano: la superficie dell'acciaio così ottenuta non presentava alcuna venatura.

Acciai ad altissimo carbonio di questo tipo, privi di damascatura, sono più resistenti e duttili, a temperatura ordinaria, degli acciai normalmente impiegati nell'industria automobilistica. Sono inoltre superplastici (si comportano cioè come melassa o vetro allo stato pastoso) nell'intervallo di temperature comprese tra 600 e 800 gradi. Grazie a questa proprietà tali acciai possono venire modellati con precisione in pezzi di forma complicata, come per esempio ingranaggi, con modestissimo ricorso a costose lavorazioni meccaniche e con processi facilmente adattabili alle esigenze della produzione in serie. Sembra quindi possibile un'estesa applicazione industriale degli acciai ad altissimo carbonio.

Non siamo i primi a vantare la riscoperta dell'arte perduta di fabbricare acciaio di Damasco. Accanto a Bréant e a Faraday va ricordato l'ingegnere russo Pavel P. Anosoff, che nel 1841 pubblicò una monografia in due volumi intitolata *Il bulat*. Anosoff era talmente entusiasta della sua scoperta che proclamò: «I nostri militi presto avranno lame di bulat, i nostri braccianti scaveranno la terra con vomeri di bulat, i nostri artigiani useranno utensili di bulat, il bulat rimpiazzerà ogni acciaio finora impiegato per realizzare oggetti destinati a mantenere una straordinaria affilatura accompagnata da un'elevata resistenza.»

Le sue previsioni non si sono realizzate. La potenzialità enorme degli acciai ad altissimo carbonio rimane a tutt'oggi inesplorata. Quantunque il nostro ottimismo sia più misurato, al pari di Anosoff riteniamo che la situazione muterà e che il segreto dell'acciaio di Damasco diverrà patrimonio dell'industria. Come dice un vecchio proverbio russo, «Il meglio del nuovo è spesso il passato a lungo dimenticato.»

Gli acciai ad altissimo carbonio possono presentare resistenza e duttilità superiori rispetto agli acciai convenzionali. La resistenza a trazione di un metallo indica il massimo sforzo che una provetta standard, ottenuta dal metallo stesso, può sopportare prima di rompersi. L'allungamento è un indice di duttilità, ossia esprime quanta deformazione il metallo può sopportare prima che avvenga la rottura. Quando gli acciai ad altissimo carbonio (*a*), per ottenere una struttura ultrafine, vengono laminati durante il raffreddamento, acquistano una maggior resistenza (per una data duttilità) rispetto agli acciai convenzionali a basso carbonio (*b*) e a quelli legati, ad alta resistenza, contenenti piccole quantità di elementi speciali (*c*). Questi tipi di acciaio presentano caratteristiche anche superiori a quelle di acciai di recente produzione (*d*).

Come il tripanosoma modifica la propria superficie

Questo parassita, il quale provoca carestie di carne e latte in buona parte dell'Africa, sfugge al sistema immunitario dell'ospite attivando sempre nuovi geni, che codificano per nuovi antigeni di superficie

di John E. Donelson e Mervyn J. Turner

Il tripanosoma è un animale microscopico, un protozoo, che vive in Africa e trascorre parte del suo ciclo vitale come parassita nel sangue dell'uomo e di altri mammiferi. In questi è causa di una malattia neurologica letale, la tripanosomiasi, il cui stadio finale nei pazienti umani è la malattia del sonno. La tripanosomiasi è endemica in un'ampia porzione dell'Africa, definita dall'area di distribuzione della mosca tse-tse, l'ospite intermedio che veicola il tripanosoma da un mammifero all'altro. Circa 50 milioni di persone corrono direttamente il rischio di contrarre la malattia e ogni anno vengono segnalati all'incirca 20 000 nuovi casi, mentre migliaia di altri casi rimangono senza dubbio ignorati. Ma, ancora più importante della minaccia diretta per gli esseri umani, è il fatto che anche il bestiame è suscettibile alla tripanosomiasi. Il tripanosoma e la mosca tse-tse riescono a rendere inabitabili per la maggior parte delle razze di bovini da latte e da carne oltre 10 milioni di chilometri quadrati di territorio africano, un'area complessivamente più estesa della superficie degli Stati Uniti. La maggior parte della popolazione umana di quel continente, avendo disponibilità di latte e di carne, scarsa o nulla, è malnutrita e contrae con facilità altre malattie.

La chiave per spiegare il successo del tripanosoma è nella sua capacità di eludere l'azione del sistema immunitario dei mammiferi. Normalmente un mammifero si difende dai virus, dai batteri o da protozoi come il tripanosoma sintetizzando anticorpi specifici diretti contro gli antigeni, o molecole «non proprie», che riconosce sulla superficie di un organismo estraneo. Gli anticorpi si legano agli antigeni e neutralizzano o uccidono l'organismo invasore. Alcune cellule produttrici di anticorpi persistono nel circolo sanguigno e conferiscono così un'immunità duratura. Questo tipo di immunità può anche essere suscitata da un vaccino che imita un'infezione naturale.

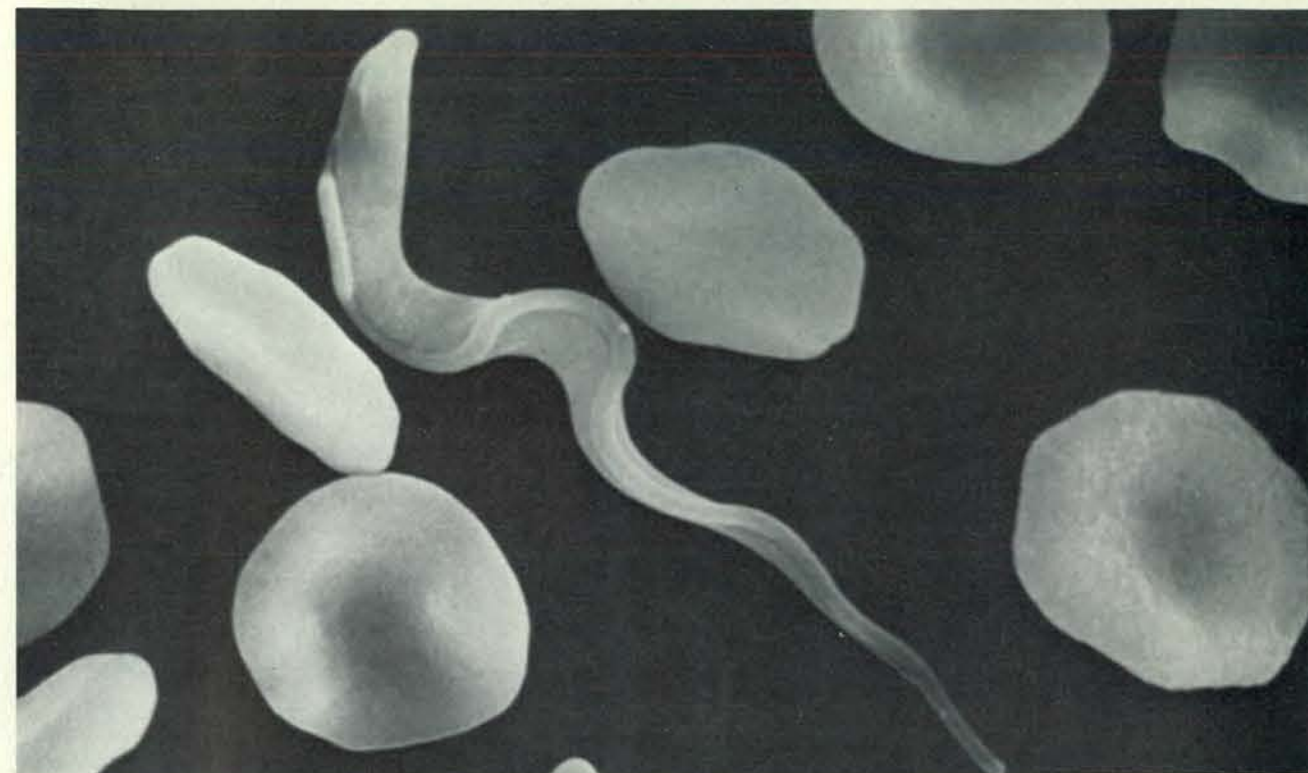
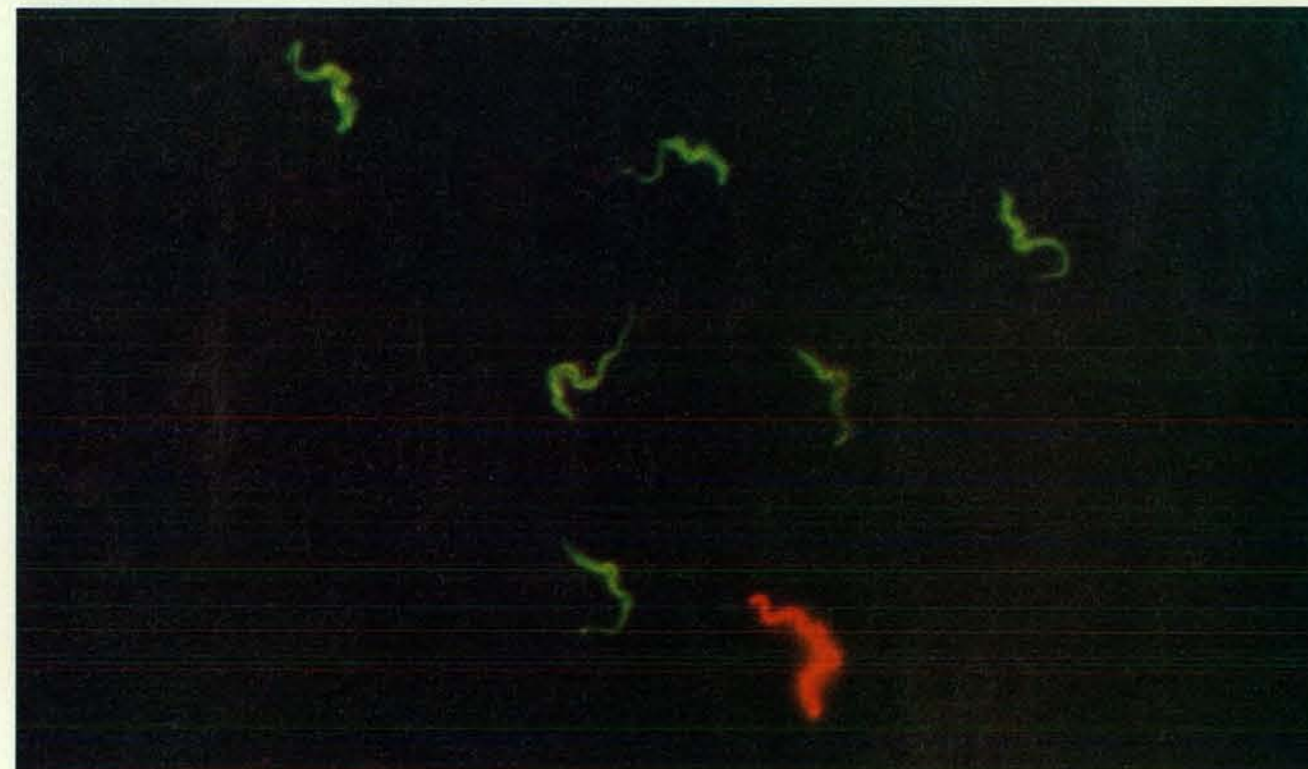
Né la risposta immunitaria all'infezione né la vaccinazione possono proteggere dall'infezione tripanosomica. Anche se i tripanosomi sono continuamente esposti, nel circolo sanguigno, al sistema immunitario del mammifero ospite (diversamente dal parassita della malaria, che trascorre la maggior parte del ciclo vitale recluso nelle cellule), hanno messo a punto un meccanismo per sfuggire alle difese dell'ospite: continuano a modificare gli antigeni che costituiscono il loro rivestimento superficiale. Nel momento in cui il sistema immunitario ha prodotto nuovi anticorpi per legarsi ai nuovi antigeni, alcuni tripanosomi hanno già perso lo strato esterno e lo hanno sostituito con un altro antigenicamente ben distinto dal primo. Il sistema immunitario dell'ospite, così sovraccaricato di lavoro, è incapace di far fronte all'infezione e i parassiti possono pertanto proliferare.

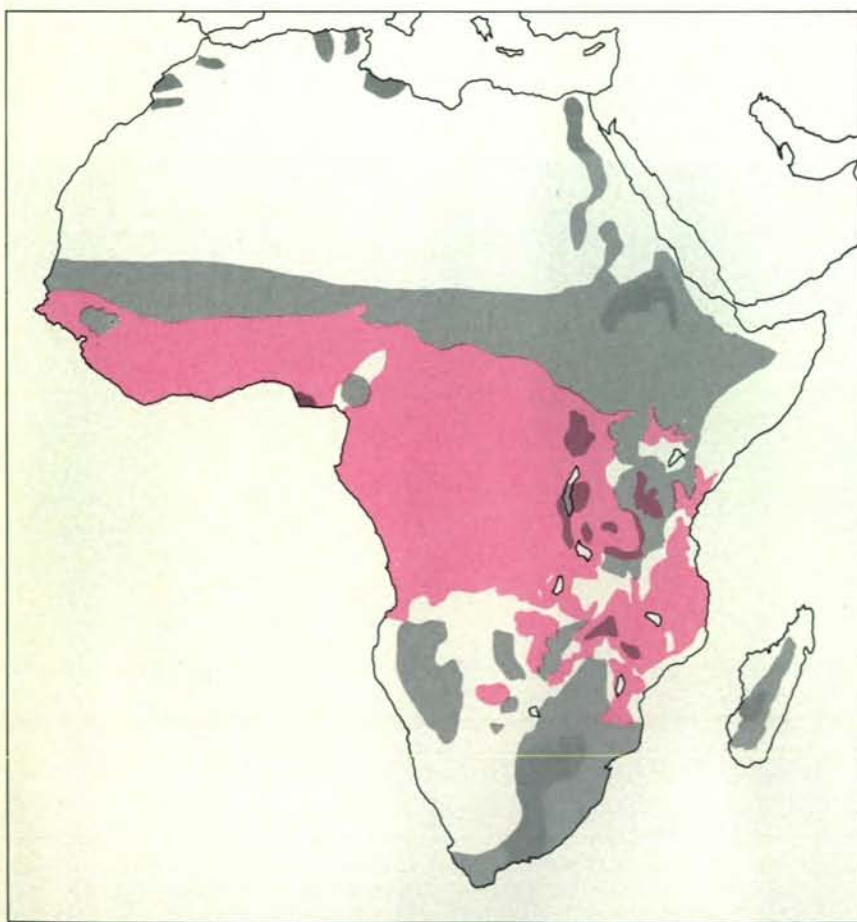
La base molecolare di questa importante variazione antigenica viene oggi profondamente studiata in numerosi laboratori in Africa, Europa e Stati Uniti. Tutto ciò che si è appreso finora fa pensare che davvero non esista un mezzo per aiutare il sistema immunitario dei mammiferi a far fronte ai tripanosomi, una volta che questi parassiti si sono stabiliti nel loro circolo sanguigno, ma che possano esservi altri mezzi di prevenzione o di cura.

I parassiti che hanno sviluppato questa strategia di difesa molto efficace sono protozoi unicellulari di lunghezza variabile tra 15 e 30 millesimi di millimetro. Vi sono numerose specie, classificate in base alla morfologia e agli ospiti che infettano. Le due importanti specie che infettano gli esseri umani sono *Trypanosoma rhodesiense* e *Trypanosoma gambiense* (così chiamate dai territori nei quali vennero identificate per la prima volta). Tre specie sono importanti parassiti del bestiame: *T. congolense, T. vivax* e *T. brucei.* Dato che si sviluppa facilmente negli animali da laboratorio e non sopravvive, invece, nel sangue umano, la terza specie è quella utilizzata preferenzialmente per la ricerca.

I tripanosomi, come molti altri parassiti, assumono forme diverse in differenti stadi del loro complesso ciclo vitale. Ingeriti da una mosca tse-tse assieme a sangue infetto di mammifero, sono accolti nello stomaco dell'insetto e qui cominciano a subire una serie di cambiamenti biochimici e strutturali; nel corso del processo perdono lo strato superficiale. Dopo circa tre settimane, infine, compaiono nelle ghiandole salivari della mosca come forma metaciclica, che presenta di nuovo uno strato superficiale.

Quando la mosca punge un mammifero, i tripanosomi metaciclici penetrano nel circolo sanguigno dell'ospite, dove si differenziano rapidamente in una forma in grado di proliferare. Negli esseri umani, la malattia che ne deriva può essere acuta o cronica, a seconda della specie infettante. In ambedue le forme, colpisce in primo luogo i vasi sanguigni e le ghiandole linfatiche, provocando una febbre intermittente, esantema e gonfiore. Proprio a questo stadio ha inizio una battaglia senza posa con il sistema immunitario dell'ospite. In seguito i parassiti invadono il sistema nervoso centrale e l'infiammazione delle membrane esterne del cervello conduce a letargo, coma e, infine, a morte.

La prima indicazione che qualcosa

Il tripanosoma, che viene introdotto nel circolo sanguigno di un mammifero dalla puntura della mosca tse-tse, è un animale unicellulare (un protozoo) con un unico flagello che si estende lungo tutto un lato della cellula. (Nella fotografia al microscopio elettronico a scansione, di Steven T. Brentano dell'Università dello Iowa, si vedono il parassita e i globuli rossi a un ingrandimento di 5500 diametri.) La superficie del tripanosoma è ricoperta da glicoproteine, dette appunto di superficie e variabili (VSG), le quali sono degli antigeni: il sistema immunitario del mammifero produce anticorpi contro di esse, uccidendo così il parassita. Un tripanosoma riesce, tuttavia, a modificare il proprio strato superficiale, dando così origine a una nuova popolazione di parassiti, che riesce a eludere le difese dell'ospite, cioè i suoi anticorpi.

Nella microfotografia a doppia esposizione, ottenuta da Klaus M. Esser del Walter Reed Army Institute of Research con un microscopio a fluorescenza, è visibile la variazione antigenica. La maggior parte dei tripanosomi è verde perché l'anticorpo che si lega alla loro VSG è marcato con un colorante che nell'ultravioletto emette fluorescenza verde. Uno, invece, è rosso; infatti ha modificato il proprio strato superficiale e presenta una diversa VSG, che è riconosciuta da un'altro anticorpo legato a un colorante con fluorescenza rossa.

Nella fotografia di Edgar D. Rowton del Walter Reed Army Institute la proboscide della mosca tse-tse e l'addome rigonfio sono rossi per il sangue ingerito. La mosca, che qui funge da animale da laboratorio, è l'ospite intermedio del tripanosoma e il suo vettore: in essa il parassita si sviluppa attraverso diversi stadi prima di essere iniettato in un mammifero, l'ospite definitivo.

La mosca tse-tse e il tripanosoma, insieme, rendono inabitabile in Africa, per la maggior parte del bestiame, una superficie di potenziale pascolo che si estende per qualcosa come 10 milioni di chilometri quadrati. Non vi è praticamente sovrapposizione tra l'area di distribuzione della mosca tse-tse *(in colore)* e le regioni dove nel continente africano si allevano i bovini *(in grigio)*.

cambia nel corso di un infezione da tripanosoma si è avuta all'inizio di questo secolo. I medici avevano osservato una marcata periodicità nella temperatura dei pazienti affetti da tripanosomiasi. Nel 1910, i ricercatori inglesi Ronald Ross e David Thomson, esaminando campioni di sangue prelevati da un paziente a distanza di alcuni giorni l'uno dall'altro, notarono che le variazioni nella temperatura corporea di quel soggetto avvenivano parallelamente a un rapido aumento e a un altrettanto rapido abbassamento del numero di parassiti nel sangue. Nell'annunciare questo risultato Ross e Thomson citavano il suggerimento fatto da un medico italiano, A. Massaglia, l'anno precedente: «Le crisi tripanolitiche sono dovute alla formazione di anticorpi nel sangue. Alcuni parassiti sfuggono alla distruzione perché si abituano all'azione di questi anticorpi. Sono questi i parassiti che provocano le ricadute». Dovettero passare oltre 50 anni prima che questa prima intuizione potesse essere confermata e spiegata.

La spiegazione ha cominciato a farsi strada nel 1965, quando Keith Vickerman dell'Università di Glasgow ha descritto per la prima volta lo spesso strato superficiale che ricopre la membrana cellulare del parassita. Ben presto si è scoperto che i singoli cloni di tripanosomi (popolazioni discendenti da una singola cellula ancestrale) hanno differenti strati superficiali. Nel 1968, Richard W. F. Le Page del Molteno Institute del Medical Research Council di Cambridge, in Inghilterra, ha analizzato le proteine antigeniche superficiali isolate da numerosi cloni e ha trovato che ogni clone presenta una proteina diversa sotto l'aspetto biochimico; le differenze erano così marcate che facevano pensare che ogni antigene dovesse rappresentare l'espressione di un diverso gene. (Un gene viene espresso da una cellula quando il suo DNA è trascritto in modo da dare origine a un filamento di RNA messaggero, il quale viene in seguito tradotto in proteina.)

Verso la metà degli anni settanta, George A. M. Cross e collaboratori hanno prodotto al Molteno Institute quattro differenti cloni, partendo da una popolazione di tripanosomi che aveva infettato un animale da laboratorio. Questi ricercatori hanno dimostrato che lo strato superficiale consiste di una matrice di molecole di glicoproteina identiche (le glicoproteine sono proteine alle quali sono attaccati gruppi carboidratici) e che queste glicoproteine sono le stesse in tutti gli individui appartenenti a un singolo clone. Avendo determinato la sequenza degli amminoacidi (subunità delle proteine) all'inizio delle glicoproteine dei quattro cloni, essi hanno notato che era diversa in ogni caso. Questa osservazione ha avvalorato la proposta di Le Page, secondo la quale

Lo strato superficiale del tripanosoma, fatto di proteine VSG, appare scuro ed esteso nella microfotografia elettronica di Laurence Tetley e Keith Vickerman dell'Università di Glasgow. La sezione trasversale del corpo e del flagello del parassita è ingrandita di 190 000 diametri. La doppia membrana appena all'interno dello strato superficiale è la membrana cellulare.

differenti antigeni di superficie devono essere codificati da geni differenti. Questi antigeni vengono attualmente chiamati glicoproteine di superficie variabili o VSG.

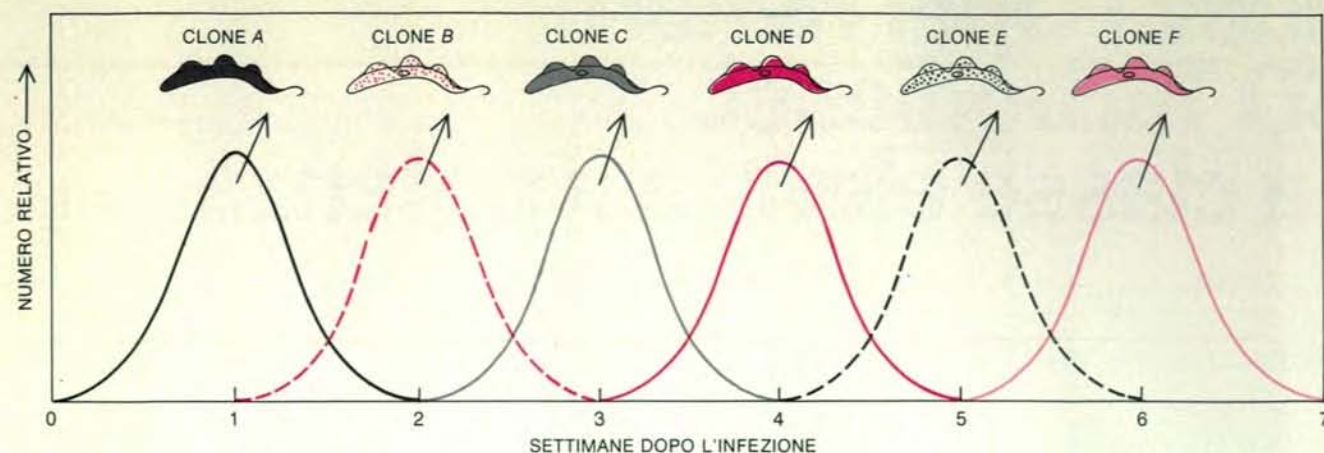
In una fase precoce del processo infettivo, il sistema immunitario genera anticorpi conformati in modo da legarsi alle particolari VSG che «vede» sullo strato superficiale dei parassiti invasori. Questi anticorpi uccidono forse il 99 per cento della popolazione parassita originaria. Tuttavia alcuni tripanosomi isolati riescono a sfuggire in quanto hanno attivato un differente gene per la VSG e sono ricoperti da un nuovo strato di molecole di VSG, alle quali non riescono a legarsi gli anticorpi presenti. Questi individui varianti danno origine a una nuova popolazione, che esprime il nuovo insieme di VSG; la nuova popolazione cresce mentre il sistema immunitario produce un altro insieme di anticorpi, i quali alla fine riusciranno a uccidere di nuovo qualcosa come il 99 per cento dei parassiti. A questo punto, però, alcuni parassiti hanno modificato di nuovo il proprio rivestimento e, pertanto, prolifererà una nuova popolazione. L'infezione procede così fintantoché l'ospite muore. Le prove di cui disponiamo suggeriscono che il passaggio da una glicoproteina di superficie variabile a un'altra abbia luogo spontaneamente. Il sistema immunitario dell'ospite non induce il passaggio, ma seleziona (con la sua incapacità a far fronte rapidamente a un nuovo antigene particolare) una variante che dà inizio a una nuova popolazione.

Il repertorio complessivo potenziale di VSG di un tripanosoma non è noto. In esperimenti controllati, i tripanosomi derivati da un'unica cellula genitrice hanno prodotto più di 100 VSG diverse, senza che vi fosse da parte loro alcuna indicazione che il corredo di geni per le VSG si fosse esaurito. È stato stimato di recente che un singolo organismo possiede almeno alcune centinaia e forse anche circa 1000 geni per le VSG. (Questo significa che dal 5 al 10 per cento della capacità genetica complessiva del parassita è impegnata nella variazione antigenica.) In natura, il pool genico (o fondo genico comune) di tutti i tripanosomi probabilmente dà l'informazione genetica per produrre un numero praticamente infinito di VSG distinte come antigeni.

Qual è la struttura delle VSG e come si attaccano queste glicoproteine alla membrana cellulare? Come avviene che in un dato momento il parassita riesca a esprimere un gene per una VSG (uno solo) partendo da un repertorio di centinaia di geni di questo tipo? Per poter conoscere qualcosa sulle VSG e su come vengano presentate sequenzialmente dal tripanosoma, numerosi gruppi di ricercatori stanno sfruttando procedimenti basati sulla tecnologia del DNA ricombinante.

Inizialmente viene isolato l'RNA messaggero (m-RNA) da un clone di tripanosomi. L'RNA messaggero viene copiato per produrre DNA complementare (c-DNA), il quale si combina con un DNA vettore e viene introdotto nei batteri. Manipolando i batteri che ospitano il DNA ricombinante, è possibile isolare alla fine una molecola di c-DNA che è, in effetti, una copia del gene per la VSG che il clone di tripanosomi sta esprimendo. Determinando la sequenza di nucleotidi (componenti del DNA) nel c-DNA e traducendola secondo il codice genetico, si può prevedere la sequenza di amminoacidi della VSG che il gene codifica.

Sequenze nucleotidiche parziali o complete sono state determinate per qualcosa come 15 c-DNA e l'analisi delle sequenze amminoacidiche dedotte rivela che ogni nuova VSG consta di circa 500 amminoacidi. I primi 20 o 30 amminoacidi costituiscono un peptide (corta catena proteica) segnale, la cui funzione consiste nel favorire lo spostamento della VSG nascente attraverso la membrana cellulare del tripanosoma. Il confronto tra la sequenza amminoacidica prevista dal c-DNA e la reale sequenza di alcune VSG ha dimostrato che, nel processo citato, il peptide segnale viene staccato. La sequenza dei successivi 360 amminoacidi è diversa nella maggior parte delle VSG e questa regione variabile è presumibilmente responsabile della diversità antigenica del parassita. Gli ultimi 120 amminoacidi (all'estremità *C*-terminale della proteina) sono del tutto analoghi in VSG diverse; sulla base del grado di somiglianza in queste regioni omologhe, le VSG possono venir classificate in due gruppi omologhi.

Normalmente sono le sequenze peptidiche nella regione *C*-terminale di una proteina di superficie ad ancorare la proteina alla membrana cellulare, ma la VSG è diversa. Nella VSG matura, gli ultimi 20 amminoacidi o poco più della regione vengono staccati e sostituiti da una struttura che contiene un oligosaccaride insolito (molecola di zucchero complessa), il quale svolge un ruolo nell'ancoraggio. Essa sembra molto simile in tutte le VSG, indipendentemente dalla sequenza della regione variabile, dato che gli anticorpi formati contro un oligosaccaride si legano in pratica a tutte le molecole di VSG. Per quale motivo questo determinante a reazione incrociata, come viene chiamato, non induce immunità naturale e perché non può servire come base per un vaccino? La ragione è essenzialmente questa: le VSG sono

Tipiche della tripanosomiasi sono le ondate successive di proliferazione del parassita nel sangue, le quali sono la conseguenza di una variazione antigenica. Una popolazione di parassiti, alcuni dei quali contengono sulla loro superficie un particolare antigene, la VSG *A*, prolifera nel circolo sanguigno per alcuni giorni. Il sistema immunitario dell'ospite produce allora anticorpi contro questa popolazione di antigeni, uccidendo la maggior parte dei parassiti. Pochi parassiti singoli sopravvivono esprimendo nuovi geni per le VSG, i quali dirigono la sintesi di nuovi antigeni, per esempio la VSG *B*. Questi parassiti danno origine a una nuova popolazione che si sviluppa fino a quando il sistema immunitario dell'ospite riuscirà a produrre nuovi anticorpi contro i nuovi antigeni. Il ciclo si ripete molte volte nel corso di una infezione cronica poiché i parassiti continuano a esprimere nuovi geni e a esibire nuove VSG. Da ogni successiva popolazione è possibile isolare singoli tripanosomi, dai quali si possono far sviluppare cloni che esprimono particolari glicoproteine di superficie variabili.

ammassate nello strato superficiale in maniera tale che il sistema immunitario si trova di fronte la regione variabile della proteina, mentre la regione omologa, comprendente il determinante a reazione incrociata, non è esposto in superficie.

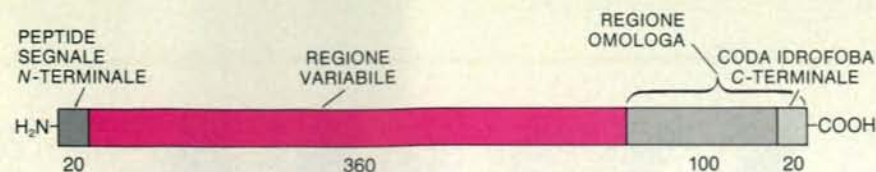
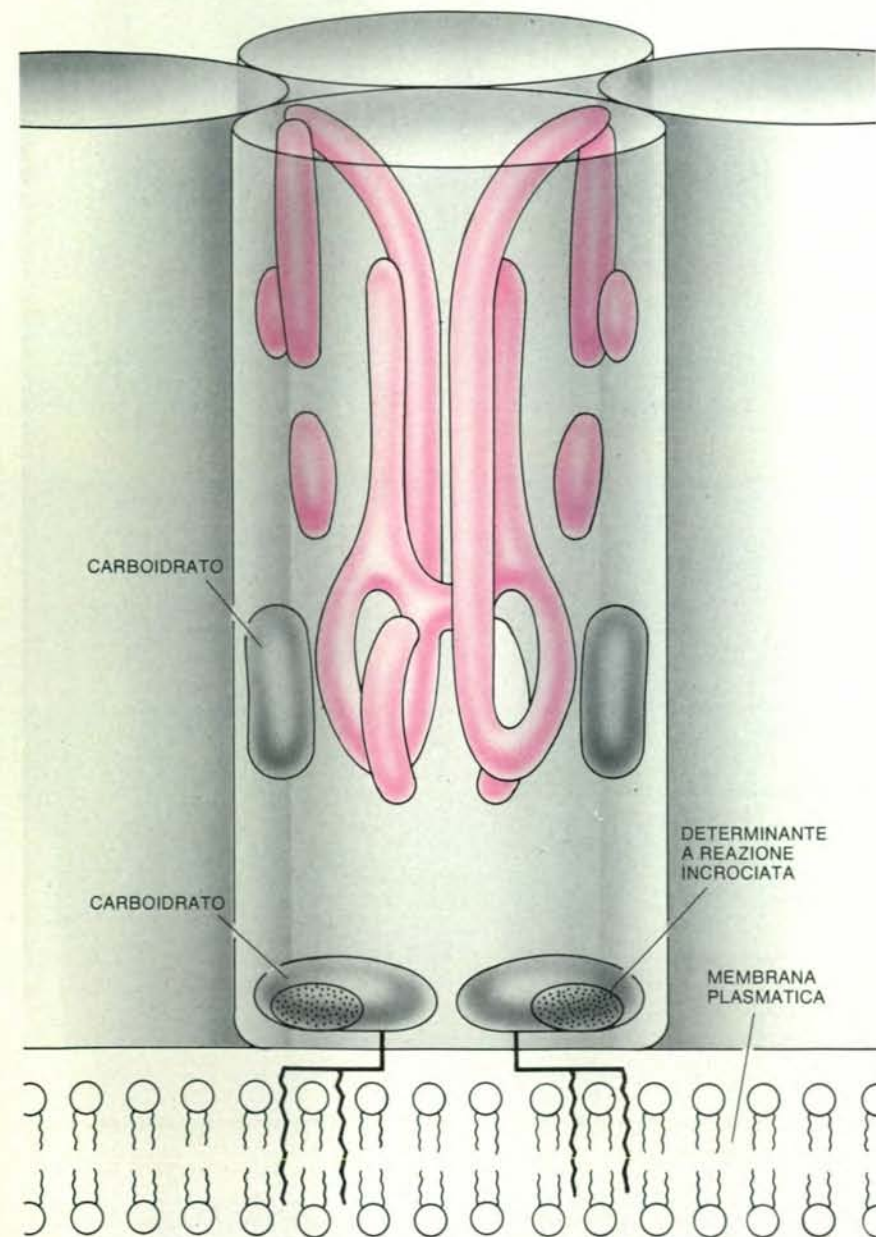
Il determinante a reazione incrociata risulta un componente di una molecola oligosaccaridica più grande, unita a sua volta a un fosfogliceride, che ha due catene di acidi grassi. Sono proprio questi ultimi a penetrare nella membrana cellulare e a mantenere al suo posto la VSG. Perché il tripanosoma sostituirebbe questa complicata struttura al normale ancoraggio *C*-terminale, nella membrana cellulare? La ragione potrebbe essere che per il parassita è di fondamentale importanza la capacità di liberarsi rapidamente dallo strato superficiale. Il tripanosoma possiede un enzima, che può scindere il legame con gli acidi grassi, liberando così la VSG dalla membrana. Il fatto che la stessa molecola specializzata di ancoraggio, soggetta a scissione catalizzata dallo stesso enzima, sia presente su tutte le VSG indipendentemente dalla loro esatta sequenza, costituisce un meccanismo rapido e universale per togliersi un involucro e sostituirlo con un altro.

Anche se il confronto tra sequenze di c-DNA ha rivelato estese diversità tra le VSG, è risultato chiaro che pochissimi cambiamenti a carico degli amminoacidi sono sufficienti a generare VSG distinte come antigeni. Presumibilmente questi cambiamenti hanno luogo in corrispondenza di particolari siti antigenici, nell'ambito di una regione variabile costituita da 360 amminoacidi. Per trovare questi siti sarà necessario conoscere nei particolari la struttura tridimensionale di questa regione. Un certo progresso verso questa meta si è avuto di recente con gli studi di cristallografia a raggi X, effettuati da Don C. Wiley, Douglas M. Freymann e Peter Metcalf della Harvard University, in collaborazione con uno di noi (Turner) del Molteno Institute.

Cinque VSG sono state cristallizzate finora, ma la maggior parte dei progressi sono stati compiuti determinando la struttura della regione variabile ottenuta da una di esse. La risoluzione raggiunta finora è sufficiente a rivelare le parti di una catena proteica che sono ripiegate in una conformazione definita elica alfa, la quale ricorda un cilindro. Risulta che metà della regione variabile sia in questa forma.

La regione variabile cristallizza come dimero (una molecola doppia costituita da due monomeri) e, in effetti, sembra che anche le VSG si aggreghino come dimeri nello strato superficiale di un tripanosoma vivente. La cristallografia dimostra che il dimero (o perlomeno la metà di esso che è stata risolta) è un fascio di eliche alfa, il cui nucleo consiste di due strutture a forma di forcina, una per ogni monomero componente. A un'estremità del dimero il nucleo interagisce con altre due eliche e forma un fascio a sei eliche; all'altra estremità i monomeri divergono in una certa misura e formano una «testa» ben distinta. Questa struttura marcatamente simmetrica deve formare l'impalcatura della rimanente metà della sequenza della regione variabile, finora non rivelata perché non è un'elica alfa. Non sappiamo ancora come sia orientata nella membrana una simile struttura di sostegno (cioè quale estremità sia «fuori») e pertanto non possiamo cominciare ad arguire dove si troveranno, sulla struttura, i principali siti antigenici.

Che cosa succede quando il tripanosoma attiva un gene per una VSG dopo l'altro, sintetizzando sempre solo una glicoproteina di superficie, distinta come antigene, alla volta? Gli esperimenti progettati per dare una risposta a questo interrogativo dipendono ancora una volta dai c-DNA del tipo descritto sopra. Ora, ciascuno di questi c-DNA, che è in effetti un gene artificiale per una VSG, funge da sonda con la quale poter localizzare copie dello stesso gene dovunque possano essere nel genoma del tripanosoma (il genoma è il contingente completo di materiale genetico).

Tutto il DNA del genoma viene digerito con un enzima di restrizione, che scinde il DNA in corrispondenza di un sito specifico all'interno di una particolare sequenza di nucleotidi. Ne risulta che il DNA genomico viene scisso in un grande numero di piccoli frammenti, ciascuno di dimensioni leggermente diverse. Questi frammenti, separati in base alla dimensione dalla elettroforesi su gel, vengono trasferiti su carta alla nitrocellulosa, alla quale si legano strettamente. Un c-DNA, che rappresenta un gene per una VSG, marcato con un isotopo radioattivo, viene posto su quella carta. Esso si ibrida con (si lega a) qualunque sequenza simile trovi sulla carta in mezzo ai frammenti prodotti dall'enzima di restrizione. La parte che non si lega viene eliminata e l'autoradiografia rivela i siti dove è avvenuta l'ibridazione. Si può così determinare quante copie diverse vi sono, in un determinato clone di tripanosomi, del gene rappresentato dalla sonda c-DNA e se,

La catena polipeptidica di una tipica VSG si compone di circa 500 amminoacidi. I primi 20 circa, in corrispondenza della estremità definita N-terminale della proteina *(a sinistra)*, costituiscono un peptide segnale, che viene staccato dal resto della proteina prima che la VSG si impianti nella membrana cellulare. I successivi 360 amminoacidi *(in colore)* costituiscono la regione variabile, che è diversa in ogni VSG distinta dal punto di vista antigenico. Gli ultimi 120 amminoacidi in corrispondenza dell'estremità C-terminale sono simili in ciascuno dei due «gruppi omologhi» delle VSG. Gli ultimi 20 amminoacidi di questa regione vengono staccati dalla catena e sostituiti da una grossa molecola che àncora la VSG alla membrana cellulare.

Le VSG dello strato superficiale potrebbero essere riunite. La struttura di parte della regione variabile di una VSG *(in colore)* si basa su dati ottenuti mediante cristallografia a raggi X *(si veda l'illustrazione a pagina 70)*; la completa estensione della VSG e la collocazione delle molecole vicine sono indicate dai cilindri in grigio. La regione variabile è un dimero, o doppia molecola, che sembra un fascio delle strutture proteiche chiamate eliche alfa. Molecole di carboidrati fiancheggiano il fascio; due altri carboidrati alla base della VSG possono incorporare una piccola molecola di zucchero; si tratta del determinante a reazione incrociata. Gli acidi grassi che si estendono da questi due ultimi carboidrati ancorano la VSG alla membrana.

in differenti cloni, il gene si trovi o no in una diversa parte del genoma.

In questo modo Piet Borst e collaboratori dell'Università di Amsterdam e Cross e collaboratori dei Wellcome Research Laboratories in Inghilterra hanno dimostrato che quando i geni per alcune VSG vengono espressi, una copia supplementare del gene è presente nel genoma: essi l'hanno chiamata «copia associata all'espressione». Etienne Pays e Maurice Steinert della Université Libre di Bruxelles sono riusciti in seguito a dimostrare che l'm-RNA per il gene di una VSG espresso viene in realtà trascritto da questa copia e non dal gene originale (copia base) che le ha dato origine. Ulteriori analisi hanno rivelato che una copia associata all'espressione si trova sempre in un particolare tipo di sito: vicino a un telomero (l'estremità di un cromosoma). In altre parole, un singolo gene per una VSG, in tutto il repertorio di geni di questo tipo, si esprime quando si duplica e viene traslocato in un sito di espressione vicino a un telomero; il passaggio da una glicoproteina di superficie variabile a un'altra si effettua spesso per rimozione e degradazione di una di queste copie e per sostituzione con la copia di un altro gene.

Questo meccanismo di copiatura e traslocazione non è l'unica fonte di variazione antigenica. John R. Young, Phelix A. O. Majiwa e Richard O. Williams dell'International Laboratory for Research on Animal Diseases di Nairobi hanno scoperto un altro meccanismo. Essi hanno trovato che in alcuni casi il numero dei differenti frammenti con i quali un particolare c-DNA sonda si ibrida non si modifica quando il gene per una VSG, rappresentato dalla sonda, si esprime: non c'è una copia associata all'espressione. La sonda si ibrida, invece, con un frammento che ha dimensioni diverse in ogni clone di tripanosomi, sia nel caso che il clone stia esprimendo il gene sia in caso contrario.

Questo significa che alcuni geni per le VSG vengono espressi senza essere duplicati e traslocati. Risulta che essi si trovano già in un sito vicino a un telomero e proprio questa vicinanza spiega le diverse dimensioni dei frammenti. Nei tripanosomi e in alcuni altri organismi capita spesso che una breve sequenza di DNA si ripeta centinaia di volte vicino a un telomero. Il numero di queste «ripetizioni a tandem» tra un gene per una VSG e un telomero varia da clone a clone. Come risultato, il frammento, che porta un dato gene per una VSG legato a un telomero, può avere dimensioni diverse in differenti cloni.

Circa la metà dei geni per le VSG, studiati fino a oggi, sarebbero geni associati al telomero. Poiché nel genoma del tripanosoma vi sono centinaia di differenti geni per le VSG, questo fatto fa pensare che debbano esservi, in tale genoma, almeno alcune centinaia di

geni per le VSG associate al telomero e, quindi, centinaia di cromosomi nel nucleo di quel parassita! (Il corredo cromosomico umano è costituito da 46 cromosomi soltanto.) Dato però che il tripanosoma ha una quantità normale di DNA per un animale unicellulare, alcuni cromosomi dovrebbero essere molto piccoli, e in effetti lo sono. Un allievo di Borst, Lex H. T. van der Ploeg, utilizzando una tecnica messa a punto da David C. Schwartz e Charles R. Cantor della Columbia University è stato in grado di distinguere il DNA nucleare del tripanosoma in quattro classi: minicromosomi della lunghezza di circa 100 000 nucleotidi, altri piccoli cromosomi da due a sette volte più lunghi, molecole di medie dimensioni (della lunghezza di circa due milioni di nucleotidi) e altre molecole che sono troppo lunghe per poter essere misurate con questo metodo.

Van der Ploeg ha trovato sui cromosomi geni per le VSG appartenenti a ogni classe di dimensioni. In un caso ha scoperto il gene per la copia base su un grosso cromosoma e il gene corrispondente per la copia associata all'espressione su un cromosoma di medie dimensioni, e ciò indica che la traslocazione di una molecola duplicata può aver luogo tra cromosomi. In un altro caso, un gene associato al telomero non è stato duplicato, ma è stato espresso nel suo sito normale vicino all'estremità di un grosso cromosoma. Queste osservazioni sono servite a dare sostegno a precedenti prove, accumulate in vari laboratori e secondo le quali, nel genoma, devono esservi più siti nei quali i geni per le VSG possano essere attivati; questi geni non sono tutti traslocati nello stesso sito di espressione vicino a un particolare telomero. Pertanto, anche se la prossimità a un telomero è necessaria per l'espressione, non può però essere sufficiente. Altri fattori devono entrare in gioco per selezionare e attivare un singolo gene per una VSG in un dato organismo, allo scopo di trascriverlo, a un dato momento, in un m-RNA.

Alcuni degli eventi che partecipano a questa attivazione possono contribuire ad aumentare la diversità delle VSG. Pays, Steinert e collaboratori hanno riferito su casi in cui non dalla duplicazione di un gene per la copia base, ma dalla ricombinazione di segmenti di almeno due differenti geni associati al telomero, ciascuno dei quali codifica per una parte della risultante VSG, è stata generata una copia funzionale associata all'espressione. Se questo tipo di ricombinazione è un fenomeno generale, deve permettere al tripanosoma di produrre un numero di VSG ancora più elevato di quello che le centinaia di geni per le VSG del suo genoma possono specificare. Esso, inoltre, suggerisce una ragione per la collocazione dei siti di espressione in corrispondenza dei telomeri: c'è una particolare probabilità che segmenti di DNA fortemente ripetitivi, come le brevi ripetizioni, definite «a tandem», vicino ai telomeri, vadano incontro a una ricombinazione.

Di quando in quando una copia associata all'espressione non viene distrutta durante il passaggio all'espressione di un altro gene, ma al contrario «indugia» per un certo tempo nel proprio sito di espressione. Uno di noi (Donelson) ha dimostrato di recente, lavorando all'Università dello Iowa, che le sequenze in corrispondenza dei confini di una particolare copia associata all'espressione sono praticamente identiche a quelle che delimitano un gene associato a un telomero. Questo risultato, assieme ad analoghi risultati ottenuti da altri gruppi di ricercatori, suggerisce che alcuni geni per le VSG, associati al telomero, possano essere copie associate all'espressione, precedentemente espresse e sopravvissute al passaggio a un nuovo gene. Forse due cromosomi si scambiano regioni adiacenti ai telomeri e, pertanto, una copia associata all'espressione viene rimossa dal sito di espressione originario e posta vicino a un differente telomero, dove rimane a disposizione per essere espressa in futuro. D'altra parte, è possibile che un singolo piccolo segmento di DNA del genoma serva da elemento di controllo mobile: un intensificatore di trascrizione che può spostarsi da un telomero all'altro e far sì che differenti geni per le VSG vengano espressi. In questo caso un gene per la copia associata all'espressione potrebbe rimanere nel proprio sito di espressione, ma essere disattivato dall'allontanamento dell'elemento di controllo.

Vi è un'altra peculiarità dei tripanosomi e degli organismi a essi affini. Gli m-RNA per le VSG (e per molte, anche se non tutte, le altre proteine) cominciano sempre con la stessa sequenza specifica di 35 nucleotidi. Questa sequenza non si trova nel corrispondente gene o in qualunque altra parte del DNA vicina a quel gene, ma viene codificata da una sequenza ripetitiva di DNA, separata dal gene. Piccole molecole di RNA trascritte da questo DNA ripetitivo producono in un modo o nell'altro copie della sequenza di 35 nucleotidi per l'inizio di ogni molecola di m-RNA. Questa sequenza di 35 nucleotidi ha presumibilmente una funzione nell'espressione dei geni del tripanosoma, ma tale funzione non è ancora nota.

La struttura di circa metà di una regione variabile della VSG è stata determinata mediante cristallografia ai raggi X da Don C. Wiley, Douglas M. Freymann e Peter Metcalf della Harvard University in collaborazione con uno degli autori (Turner). La regione variabile cristallizza come un dimero. Una mappa di densità elettronica *(a sinistra)*, che è una vista perpendicolare all'asse del dimero, rivela un fascio di eliche alfa bastoncellari. Vi sono sei eliche a un'estremità (qui in basso, anche se non si conosce l'effettiva orientazione del dimero) e da quattro a sei al centro; i due monomeri divergono in alto in una struttura a due teste. Le tre sezioni *(a destra)* sono viste lungo l'asse del dimero vicine alle parti superiore, centrale e inferiore.

Riassumiamo ora tutte le conoscenze sui meccanismi di variazione antigenica nel tripanosoma. Vi sono, nel genoma di ogni organismo, centinaia di geni che codificano per le VSG. Essi possono trovarsi all'interno dei cromosomi o vicino ai loro telomeri. Solo una VSG alla volta viene trascritta, e il gene che ne è responsabile è sempre vicino a un telomero. Per essere trascritto, un gene interno (la copia base) deve essere duplicato e traslocato (come copia associata all'espressione) in uno dei molti siti di espressione, i quali si trovano tutti vicini ai telomeri. D'altra parte, un gene per la VSG, associato a un telomero, non deve essere duplicato per venire espresso (anche se, di fatto, può talvolta esserlo). La diversità antigenica può essere ulteriormente incrementata dalla ricombinazione. I meccanismi molecolari precisi che innescano il passaggio da un gene per una VSG a un altro non sono ancora noti. Sembrerebbe, tuttavia, che essi siano così complessi e vari che è sostanzialmente impossibile eluderli; per questo motivo è improbabile che si possa riuscire a mettere a punto un vaccino efficace contro i tripanosomi presenti nel circolo sanguigno.

Può darsi, invece, che si riesca a mettere a punto un vaccino contro i tripanosomi metaciclici. Come abbiamo già ricordato, lo stadio metaciclico è lo stadio di sviluppo finale del tripanosoma, che si svolge nelle ghiandole salivari della mosca tse-tse e sono appunto dei parassiti metaciclici che vengono inoculati nel circolo sanguigno di un mammifero quando una mosca lo punge. Steven L. Hajduk, J. David Barry, Keith Vickerman dell'Università di Glasgow, e Klaus M. Esser del Walter Reed Army Institute of Research di Washington hanno trovato che i tripanosomi metaciclici possono mostrare alla superficie della loro cellula solo un ridotto sottoinsieme di VSG, probabilmente soltanto 15. In collaborazione con Esser, uno di noi (Donelson) ha studiato i c-DNA di diverse VSG metacicliche e ha dimostrato che queste proteine hanno all'incirca la stessa regione omologa *C*-terminale del-

I geni per la copia base, all'interno dei cromosomi di un tripanosoma *(a sinistra)*, non vengono normalmente trascritti in RNA messaggero (m-RNA). Perché uno di essi *(C)* possa essere espresso deve venir duplicato in modo da fornire una copia associata all'espressione. Il DNA della copia viene traslocato in un sito di espressione vicino a un telomero (l'estremità di un cromosoma) e trascritto in RNA messaggero, che è poi tradotto nella proteina VSG. Altri geni per le VSG sono già vicini a un telomero *(a destra)*: si tratta dei geni associati al telomero, che possono venire espressi senza formare una copia associata all'espressione, ma di tanto in tanto anch'essi vengono duplicati e traslocati. Una copia associata all'espressione va generalmente perduta durante una variazione antigenica, ma talvolta «indugia» e diventa un gene associato a un telomero e non espresso. Stranamente, una sequenza di 35 nucleotidi a un'estremità dell'RNA messaggero non viene codificata dal gene per una VSG, ma viene specificata invece da una regione separata di DNA ripetitivo.

le VSG presenti nel circolo sanguigno; evidentemente esse si attaccano alla membrana cellulare nello stesso modo. Inoltre, i geni per le VSG metacicliche sembrano legati al telomero, come molti geni per le VSG del circolo sanguigno. Pertanto non è ancora chiaro perché i tripanosomi metaciclici non possano esprimere l'intera gamma delle VSG. Può esservi nondimeno un modo per trarre vantaggio dalla limitatezza del loro repertorio o dal loro meccanismo di commutazione e riuscire così a produrre un vaccino efficace.

I farmaci attualmente disponibili per curare la tripanosomiasi sono estremamente tossici e inoltre non riescono a impedire la reinfezione; è possibile tuttavia che si riescano a reperire nuove forme di chemioterapia. Il tripanosoma non riesce a sopravvivere in un mammifero senza lo strato superficiale. Un farmaco che interferisca con il fosfoglicerìde che àncora la VSG alla membrana cellulare, o che attiva l'enzima che libera la VSG, potrebbe pertanto essere un agente terapeutico efficace. Gli RNA messaggeri dei mammiferi non hanno l'insolita sequenza di 35 nucleotidi descritta prima: un farmaco che interferisca con la sintesi di questa sequenza potrebbe pertanto disabilitare selettivamente il tripanosoma. Si potrebbero anche trovare farmaci che agiscono contro due organelli subcellulari, apparentemente esclusivi dei tripanosomi. Uno è il glicosoma, un aggregato di enzimi legato alla membrana; l'altro è il cinetoplasto, un'appendice dell'unico grosso mitocondrio del parassita. Un farmaco che interferisca con una funzione metabolica di questi organelli, o con qualche altra via metabolica esclusiva finora sconosciuta, potrebbe uccidere il tripanosoma senza danneggiare l'ospite.

Vi sono alcuni altri possibili modi di affrontare il problema. La mosca tse-tse può essere sradicata in breve tempo in aree limitate mediante irrorazione di insetticidi oppure distribuendo nell'ambiente maschi sterili. Ma questi metodi non possono essere efficaci per una regione che si estende per qualcosa come 10 milioni di chilometri quadrati e che comprende una ventina di paesi. Sembra che alcune razze di bovini allevate da tribù nomadi abbiano sviluppato una parziale resistenza alla tripanosomiasi. Esse non producono molta carne o latte, ma può darsi che si riesca a incrociarle con razze più produttive. Un'altra via possibile sarebbe la domesticazione di animali selvatici come il taurotrago orice o l'orice, che non sembrano minacciati dai tripanosomi e che finirebbero con il sostituire i bovini domestici.

Negli ultimi anni è risultato evidente che i tripanosomi africani sono eccellenti animali da laboratorio. L'instancabile e continua ricerca sul loro strato superficiale variabile fornirà l'informazione di base su argomenti diversi quali, per esempio, il controllo dell'espressione genica, l'attacco e il funzionamento delle proteine di membrana, la struttura dei telomeri cromosomici e la loro duplicazione e, infine, i meccanismi molecolari che danno origine alla diversità biologica.

Conoscenze fondamentali di questo tipo dovrebbero a loro volta contribuire ad assolvere un compito davvero pressante riguardante la salute pubblica: il controllo della tripanosomiasi. L'immediato futuro dovrebbe mostrarci se la nuova informazione di base sui tripanosomi può davvero essere applicata al controllo della malattia e forse anche alla sua eliminazione.

La sostituzione funzionale dell'orecchio

Le protesi impiantabili, progettate per attivare direttamente il nervo acustico con stimoli elettrici, sono particolarmente promettenti in caso di sordità dovuta a danno delle cellule ciliate dell'orecchio interno

di Gerald E. Loeb

I recenti progressi dell'elettronica e della neurofisiologia hanno gettato le basi per lo sviluppo di un nuovo campo di ricerca, quello del controllo nervoso. Questa disciplina sperimentale si basa sullo scambio di informazioni tra un circuito elettronico e il sistema nervoso allo scopo di studiare o di integrare una funzione biologica. Uno dei suoi principali obiettivi è la messa a punto di dispositivi protesici per sostituire parti lese del sistema nervoso umano. Gran parte del progresso in questo campo è derivato dalla lunga collaborazione tra gruppi di studiosi impegnati nella ricerca sia di base sia applicata sotto l'egida del Neural Prosthesis Program del National Institute of Neurological and Communicative Disorders and Stroke. Il mio lavoro sulla sostituzione funzionale dell'orecchio umano è stato compiuto nell'ambito di questo progetto ed è stato eseguito in collaborazione con un gruppo di ricercatori dell'Università della California a San Francisco.

La particolare protesi nervosa che viene presa in considerazione in questo articolo è destinata a impianto cocleare in pazienti con sordità neurosensoriale. In questi pazienti il funzionamento delle cellule sensoriali ciliate della coclea, struttura a forma di chiocciola situata nella parte centrale dell'orecchio interno, risulta deteriorato. In un soggetto con udito normale il suono si propaga attraverso il meato acustico esterno fino a una membrana, il timpano, che trasmette le vibrazioni dell'aria al sistema di ossicini dell'orecchio medio. Il più interno di questi ossicini, la staffa, è a contatto con la finestra ovale, un'apertura situata alla base della coclea e coperta da una membrana, che trasmette la vibrazione al liquido che riempie l'interno della coclea; qui la vibrazione è recepita dalle cellule ciliate situate in una struttura ben definita, l'organo di Corti. Le cellule ciliate sono disposte in quattro lunghe file sulla membrana basilare (o basale), che costituisce un divisorio flessibile situato alla base dell'organo di Corti e separa due dei tre canali paralleli, ad avvolgimento elicoidale, della coclea. Le cellule ciliate trasformano la vibrazione della membrana basilare in un segnale elettrico, che si propaga lungo il nervo acustico fino al cervello.

Vari tipi di protesi cocleari sono stati progettati per sostituire questo elaborato processo di trasmissione in pazienti con sordità neurosensoriale. Tutti i dispositivi proposti hanno quattro caratteristiche in comune: un microfono per ricevere il suono, un elaboratore microelettronico per trasformare il suono in segnali elettrici, un sistema di trasmissione per inviare i segnali alle componenti impiantate e un elettrodo lungo e sottile inserito chirurgicamente nei recessi più interni della coclea in modo che il dispositivo possa attivare direttamente, con stimoli elettrici, le fibre del nervo acustico in uno o più punti.

Un dispositivo di questo tipo, un modello relativamente semplice e a canale singolo, è stato approvato di recente dalla Food and Drug Administration per l'impianto in pazienti affetti da sordità neurosensoriale. Esso è stato progettato da William House dello House Ear Institute di Los Angeles ed è prodotto dalla 3M.

Sembra che dispositivi più sofisticati, a canale singolo o a canali multipli, tuttora in fase di studio, possano fornire ai suddetti pazienti una percezione dei suoni più vicina alla realtà. Lo studio approfondito delle risposte di questi pazienti e di animali da laboratorio a vari tipi di stimolazione elettrica prodotti da questi dispositivi ha in più il vantaggio di fornire nuove intuizioni sui normali processi di codificazione e di decodificazione delle informazioni acustiche da parte del sistema nervoso.

In un certo senso la struttura dell'informazione acustica è molto semplice. Il segnale ricevuto da un microfono, o da un orecchio, può essere interamente descritto da una singola funzione di tempo, che caratterizza il movimento di un singolo punto nello spazio, per esempio il centro

Una protesi acustica sperimentale, progettata dai ricercatori dell'Università della California a San Francisco, è stata impiantata nell'orecchio, visibile in sezione, di un paziente. A destra è illustrato un ingrandimento della coclea, la struttura a forma di chiocciola situata al centro dell'orecchio interno. (Le file di cellule ciliate sensoriali, che appaiono in quest'immagine, dovrebbero mancare in un paziente con sordità neurosensoriale tipica). Questa particolare protesi cocleare è un dispositivo bipolare a otto canali che emette stimoli a otto diverse frequenze per suddividere nella coclea gruppi di fibre del nervo acustico tramite otto coppie ravvicinate di contatti elettrici, distribuiti lungo l'elettrodo impiantato. Si presuppone che dispositivi a canali multipli di questo tipo, attualmente ancora allo stadio di ricerca clinica, rendano possibile una percezione più realistica dei suoni nei pazienti con sordità neurosensoriale rispetto ai dispositivi a canale singolo ora disponibili. L'illustrazione di pagina 80 mostra il funzionamento delle principali componenti di un «elemento motore» a quattro canali per questo tipo di dispositivo. Sono presenti solo quattro canali, ma l'elettrodo (che non può essere facilmente sostituito a causa del tessuto cicatriziale) è stato progettato per riceverne otto; un raccordo accessibile chirurgicamente alla base dell'elemento motore ne facilita la sostituzione con versioni perfezionate.

del diaframma del microfono o della finestra ovale della coclea. Nonostante la sua semplicità, questo segnale contiene tipicamente un'informazione complessa riguardante la sorgente del suono; il sistema nervoso è in grado di analizzarlo per ricavarne l'informazione. Nel caso dell'uomo, questa analisi è sufficiente, in genere, per identificare la persona che parla e le parole che pronuncia. L'aggiunta di un secondo canale di informazione in un punto diverso dello spazio - l'altro orecchio - permette di distinguere sorgenti sonore multiple in base alle loro posizioni relative.

Nel corso di quest'analisi la funzione unidimensionale deve essere scissa nelle frequenze che la compongono. La suddivisione di una forma d'onda complessa di questo tipo nelle sue componenti spettrali con il metodo matematico classico dell'analisi di Fourier è oggi una pratica comune in elettronica ed è spesso ottenibile, con i moderni strumenti, con la semplice pressione di un tasto. Dalla fine del XIX secolo è stato dato per assunto che il cervello si serve di una forma analoga di analisi del segnale, almeno nei primi stadi dell'ascolto. Inoltre, si è riconosciuto che i canali attraverso i quali l'informazione viene trasmessa nel sistema nervoso, cioè i singoli neuroni, sono implicitamente molto lenti. Un sistema analitico costituito da un grande numero di canali d'informazione, in cui ogni canale può ricevere solo 300 impulsi al secondo circa, deve fornire in un modo o nell'altro un'analisi spettrale accurata e quasi istantanea dei segnali che coprono una banda di frequenza compresa tra i 20 e i 20 000 hertz (cicli al secondo). Nonostante la presenza di un notevole rumore di fondo in ogni canale, la prestazione globale del sistema non deve, inoltre, essere deteriorata oltre un intervallo dinamico di un milione a uno (120 decibel) a partire dalla soglia di udibilità fino alla comparsa del dolore.

In che modo l'orecchio realizza questo obiettivo? Hermann von Helmholtz fu uno dei numerosi fisici che, nel XIX secolo, giunsero a riconoscere che l'organo di Corti può essere interessato in qualche modo nella «campionatura» dei suoni che vengono fisicamente separati nelle loro componenti spettrali tramite la vibrazione simpatica (per risonanza) della membrana basilare. I primi studi sul movimento della membrana basilare, eseguiti da Georg von Békésy negli anni cinquanta, hanno stabilito che, quando un suono di una data frequenza è applicato alla base della coclea, provoca la vibrazione della membrana basilare con la massima ampiezza in un punto particolare, che ha una risonanza naturale per la frequenza del suono applicato. Una forma d'onda acustica complessa, composta di molte frequenze sinusoidali singole, genera una distribuzione spaziale, lungo la membrana basilare, dei punti di massima vibrazione per risonanza; in questa distribuzione le frequenze più alte si riscontrano alla base della spirale cocleare e quelle più basse all'apice. Le cellule ciliate dell'organo di Corti trasducono l'energia meccanica di movimento della membrana basilare in energia elettrica che eccita le fibre adiacenti del nervo acustico. Ne risulta che impulsi a bassa frequenza vengono generati nell'insieme parallelo di fibre del nervo acustico. Ogni fibra trasmette informazioni riguardanti solo l'ampiezza della vibrazione in un dato punto.

Questa teoria della «organizzazione tonotopica della coclea» nella percezione uditiva è alla base della progettazione di molte protesi cocleari a canali multipli, anche se, dalla trattazione che segue, risulta chiaro che non è probabilmente sufficiente a spiegare molti fenomeni psicofisici importanti nella percezione uditiva, tra cui alcuni attributi cruciali della normale percezione del linguaggio.

Dei tre componenti impiantabili messi a confronto, quello in alto è una protesi cocleare a canale singolo progettata da William House dello House Ear Institute di Los Angeles e prodotta dalla 3M. È il primo dispositivo di questo tipo ad aver ricevuto l'approvazione dalla Food and Drug Administration. I componenti impiantati consistono di un semplice filamento metallico avvolto a elica, che porta direttamente a una coppia di elettrodi di forma sferica, uno dei quali è inserito nel canale timpanico alla base della coclea. (L'altro elettrodo è «messo a terra» in una parte vicina dell'orecchio medio.) Al centro si nota un dispositivo più perfezionato, a canale singolo, progettato da Inge J. Hochmair-Desoyer e da Erwin S. Hochmair del Politecnico di Vienna e oggi in fase di sperimentazione da parte della 3M: all'interno di un contenitore di plastica trasparente si può vedere l'insieme dei circuiti del ricevitore; in questo dispositivo entrambi gli elettrodi sferici sono progettati per essere impiantati al di fuori della coclea nell'orecchio medio. In basso è riportato un dispositivo a canali multipli messo a punto da Graeme M. Clark e collaboratori all'Università di Melbourne e prodotto dalla Nucleus Limited. Il contenitore di titanio ermeticamente chiuso contiene il complesso sistema di circuiti, necessari per decodificare un segnale trasmesso, che seleziona e attiva uno dei 22 contatti situati lungo l'elettrodo.

L'impianto di una protesi cocleare si basa sul fatto che nei pazienti con sordità neurosensoriale molte fibre del nervo acustico rimangono spesso intatte. Come per la maggior parte delle cellule eccitabili elettricamente, i neuroni che rimangono funzionanti possono essere indotti a emettere una scarica di impulsi nervosi, che si propagano attivamente, mediante l'applicazione di correnti elettriche esterne caratterizzate da intensità, durata e direzione adeguate. Questi «potenziali evocati» arrivano al cervello proprio come gli impulsi generati da segnali acustici trasdotti da cellule ciliate integre e pertanto il cervello li interpreta come suoni.

L'intensità del suono percepito dipende grosso modo dal numero di fibre nervose attivate e dalla velocità con cui emettono scariche di impulsi. Entrambe queste variabili sono funzioni dell'ampiezza della corrente applicata. Il tono è correlato, invece, al punto della membrana basilare da cui queste fibre nervose originariamente hanno derivato le informazioni acustiche, in accordo con la teoria dell'organizzazione tonotopica. Con canali di stimolazione sufficientemente indipendenti, ciascuno dei quali controlli l'attività di un piccolo fascio localizzato di fibre del nervo acustico, si potrebbe teoricamente ricreare la risposta nervosa normale a stimoli acustici di composizione spettrale qualsiasi. Il cervello poi elaborerebbe l'informazione secondo i normali schemi e il soggetto, a questo punto, potrebbe «sentire» i «suoni».

Il problema che deve affrontare chi progetta una protesi nervosa è che tutte le fibre del nervo acustico sono immerse in uno stesso ambiente, costituito da tessuti e liquidi a conduzione elettrica. Una corrente elettrica immessa in questo mezzo tende così a diffondere simmetricamente dalla sorgente; ne risulta che la sua densità diminuisce con il quadrato della distanza a partire da una sorgente unipolare. Perché un tale stimolo possa attivare selettivamente un particolare fascio di fibre del nervo acustico, l'elettrodo deve essere molto più vicino a queste fibre che alle altre.

Questa limitazione è difficile da superare. Entrando nell'organo di Corti, sulla membrana basilare, le fibre del nervo acustico si disperdono ampiamente. Una schiera longitudinale di elettrodi può essere inserita nel canale timpanico e fatta passare lungo la membrana basilare a una distanza di 25 millimetri circa dalla finestra rotonda, una seconda apertura alla base della coclea. Lungo questo tratto sono distribuite le fibre che normalmente trasmettono le informazioni riguardanti i suoni con frequenze superiori a 500 hertz. La distanza minima tra le fibre del nervo acustico e la sede ottimale dell'elettrodo nel canale timpanico (lungo la parete mediale) è dell'ordine di un millimetro. Un impulso prodotto da una corrente di stimolazione pari a quattro volte la soglia, diretto alle fibre in una data posizione (e caratterizzato da un'intensità moderata), influenzerà tutte le fibre che si trovano nel raggio di due millimetri in entrambe le direzioni.

Quando le variazioni naturali nelle dimensioni delle fibre e la suscettibilità alla stimolazione vengono analizzate nelle loro componenti, tale influenza risulterà ancora più dispersa. Ovviamente solo un piccolo numero di siti indipendenti di stimolazione con un qualsiasi intervallo dinamico apprezzabile può essere contenuto nello spazio disponibile. Per la percezione del linguaggio la banda di frequenza critica è compresa tra i 500 e i 3000 hertz, il che rappresenta un tratto della membrana basilare inferiore a 14 millimetri.

Questo tratto può a sua volta contenere forse due o tre siti di stimolazione indipendenti, comprendenti un intervallo dinamico ragionevole. D'altro canto, gli esperimenti in cui il linguaggio viene simulato tramite un piccolo numero di generatori di toni singoli a modulazione di ampiezza presuppongono l'esistenza di almeno sei canali per raggiungere l'intelligibilità.

Un modo per ripartire efficacemente le fibre in un numero adeguato di canali distinti è quello di applicare una stimolazione bipolare. Ricorrendo a questo sistema i poli positivo e negativo dell'impulso di corrente risultano l'uno vicino all'altro. (Nella stimolazione unipolare il polo negativo necessario per completare il circuito è un contatto ampio e remoto che funge da presa di terra per tutte le sorgenti.) Le linee di densità di corrente attorno a una coppia di elettrodi bipolari sembrano le linee di un campo magnetico attorno a una calamita: esse hanno una forma ellittica e il loro asse maggiore è orientato parallelamente alla linea che congiunge i due poli *(si veda l'illustrazione a pagina 81)*.

Poiché l'attivazione dei neuroni richiede l'induzione di correnti che scorrono longitudinalmente lungo i loro prolungamenti, elettrodi bipolari orientati ad angolo retto rispetto all'asse lungo del canale timpanico (radialmente all'elica) possono attivare selettivamente solo le fibre nervose locali che passano immediatamente al di sopra di essi. Il gradiente di densità di corrente nei tessuti, che si estende verso l'esterno a partire da una coppia di elettrodi bipolari, è inoltre molto più ripido di quello determinato da un elettrodo unipolare, ottenendosi così una riduzione della diffusione dello stimolo anche ad alte intensità. Esperimenti sia su soggetti umani sia su animali entrambi con protesi cocleari, hanno dimostrato che questi elettrodi bipolari radiali producono un'attivazione più localizzata su intervalli dinamici più ampi rispetto alle configurazioni unipolari o bipolari longitudinali.

Il numero reale di canali indipendenti che si riescono a stabilire e i loro intervalli dinamici dipendono dalle condizioni delle fibre del nervo acustico rimaste vitali. Queste a volte vanno incontro a una degenerazione retrograda a una certa distanza dall'organo di Corti e, di conseguenza, si riduce la loro vicinanza alla schiera di elettrodi posti nel canale timpanico. In alcuni pazienti, tuttavia, sembra che queste coppie di elettrodi bipolari radiali possano essere posizionate a circa due millimetri l'una dall'altra senza un'interazione significativa. Questa spaziatura consente di disporre otto canali indipendenti nell'area corrispondente alle frequenze vocali, compresa tra 10 e 24 millimetri a partire dalla finestra rotonda.

Esistono vari problemi tecnici nel posizionamento di questi otto canali bipolari di stimolazione del nervo acustico nell'uomo. I primi 24 millimetri della membrana basilare sono avvolti nel primo giro e mezzo dei due giri e mezzo che costituiscono la spirale cocleare. La finestra rotonda nel canale timpanico è situata dietro la cavità dell'orecchio medio, obbligando il chirurgo a lavorare attraverso il lungo e stretto meato acustico esterno. Questo significa che l'elettrodo deve essere rettilineo mentre viene immesso nella coclea. Tuttavia, non può essere semplicemente spinto nella spirale cocleare come se fosse il tubo di gomma che l'idraulico fa passare attraverso un tubo di scarico ricurvo. La forma semicilindrica della sezione trasversale del canale timpanico fa sì che un oggetto flessibile che prema contro la parete laterale venga deviato verso l'alto contro la membrana basilare, che è estremamente fragile. Il liquido nel canale cocleare situato proprio al di sopra della membrana ha un elevato contenuto di potassio, il che lo rende tossico per le fibre del nervo acustico se penetra nel canale timpanico.

La teoria della «organizzazione tonotopica della coclea» spiega in che modo il sistema cocleare di rilevazione dei suoni risolve una forma d'onda acustica complessa nelle sue componenti spettrali. In questa visualizzazione della teoria, un'onda sonora composta che si propaga nel liquido cocleare provoca una vibrazione simpatica della membrana basilare, il sottile e flessibile divisorio che separa due dei tre canali della coclea ad avvolgimento elicoidale. I massimi della vibrazione sono distribuiti nello spazio lungo la membrana; le frequenze più alte sono rilevate alla base della coclea e le più basse all'apice.

Una soluzione è stata quella di utilizzare un elettrodo molto sottile e flessibile, che teoricamente può essere posizionato con un trauma minimo. Tuttavia è difficile, se non impossibile, controllare l'esatta orientazione dei poli di stimolazione dell'elettrodo; esso tende a posarsi contro la parete esterna, lontano dalle fibre nervose che deve attivare. La selettività e l'intervallo dinamico di una schiera di elettrodi bipolari radiali dipende proprio dall'accurato posizionamento di ogni coppia sulla parete mediale.

Questo obiettivo può essere raggiunto solo usando un elettrodo relativamente spesso, dotato di «memoria» per una forma elicoidale che lo fa aderire alla parete mediale lungo i vari giri della coclea. Queste proprietà sono state realizzate in una schiera di otto coppie di elettrodi bipolari progettata dal gruppo di ricerca dell'Università della California a San Francisco. Il dispositivo unisce le proprietà meccaniche di un elemento portante in gomma e silicone a quelle dei fili di platino-iridio, piatti e sovrapposti, in uscita dai 16 contatti degli elettrodi distribuiti lungo la spirale *(si veda l'illustrazione alle pagine 76 e 77)*.

Anche il trasferimento della corrente di stimolazione dai circuiti elettronici ai tessuti biologici è una questione delicata. Le correnti elettriche nei conduttori metallici sono trasportate dagli elettroni, mentre nelle soluzioni acquose che

circolano nell'organismo sono trasportate dagli ioni. Le reazioni elettrochimiche all'interfaccia dei metalli anche maggiormente biocompatibili e la complessità dei liquidi corporei sono fonte di rischio sia per gli elettrodi sia per i tessuti. La corrosione elettrolitica di molti metalli supposti inerti aumenta rapidamente per l'azione chelante degli ioni cloro, i più importanti ioni negativi dei liquidi extracellulari. Questo processo modifica le proprietà elettriche dell'interfaccia e provoca la liberazione di ioni di metalli pesanti molto tossici nelle immediate vicinanze dei neuroni, che sono tra le cellule corporee più sensibili. Anche la semplice elettrolisi dell'acqua può produrre bolle di idrogeno e di ossigeno con azione meccanica dirompente, oltre a determinare concentrazioni locali di ossidrili e di ioni idronio, che possono raggiungere livelli tossici rispettivamente di alcalinità e di acidità.

Solo negli ultimissimi anni queste reazioni e le loro conseguenze sono state studiate a livello quantitativo, in gran parte da S. Barry Brummer e collaboratori agli EIC Laboratories, Inc., di Norwood, nel Massachusetts. Si è così stabilito che determinate condizioni piuttosto rigorose, in cui le correnti elettriche attraversano particolari metalli quali il platino e l'iridio, possono indurre senza rischi correnti ioniche nei liquidi corporei tramite reazioni elettrochimiche reversibili e non tossiche.

Dopo il posizionamento degli elettrodi e l'applicazione della corrente di stimolazione, che cosa sente il paziente? In genere, finora, i risultati sono stati incoraggianti ma sono ancora un misto di previsto e imprevisto.

Il campione di pazienti di gran lunga più numeroso è stato esaminato con stimolatori a canale singolo. Forme d'onda moderatamente filtrate, provenienti direttamente da un microfono, vengono trasmesse a uno stimolatore impiantato, che attiva alcune fibre del nervo acustico presso la base della coclea. Presumibilmente le sensazioni che si ottengono sono quelle di un rumore complesso, modulato in ampiezza. Vi è solo la generica percezione di tono e solo per frequenze di stimolazione inferiori a qualche centinaio di hertz. Invece la percezione del ritmo e dell'intensità del suono udito è notevole. Nei pazienti affetti da sordità profonda e per i quali anche le più potenti

In una protesi a quattro canali il suono è captato da un microfono da risvolto e inviato per l'elaborazione a una scatoletta che il paziente porta addosso, munita di controlli di tono e di intensità. Un circuito di controllo automatico dell'amplificazione nel chip microelettronico contenuto nella scatoletta riduce innanzitutto le ampie fluttuazioni di intensità che caratterizzano il suono in un ambiente normale a un intervallo molto più limitato, adatto per la stimolazione elettrica dell'orecchio interno. Quindi una serie di filtri passa-banda divide la complessa forma d'onda della corrente alternata in quattro canali di frequenza che corrispondono alle principali formanti prodotte dal sistema vocale umano nel realizzare i suoni delle vocali. I segnali a banda stretta che ne risultano (uno dei quali, in questo esempio, è zero) modulano l'ampiezza di quattro oscillatori indipendenti a radiofrequenza, montati sul complesso bobina-antenna. Le antenne trasmettenti sono fissate sul cuoio capelluto per mezzo di coppie di magneti ceramici in punti corrispondenti ai solenoidi riceventi impiantati. Ogni circuito ricevente agisce come una unità radioricevente indipendente in modulazione di ampiezza, che serve a demodulare la frequenza portante e a riottenere il segnale originario a banda stretta, trasmesso direttamente a una coppia di contatti di stimolazione sull'elettrodo inserito nel canale timpanico. Questo schema è limitato a un numero esiguo di canali. Un gruppo di ricercatori della Stanford University ha messo a punto di recente un sistema perfezionato, a otto canali, nel quale le caratteristiche particolareggiate del segnale in uscita di ciascun canale sono digitalizzate per essere trasmesse su un'unica frequenza portante e decodificate nei circuiti del ricevitore impiantato.

protesi acustiche convenzionali non sono efficaci, queste informazioni sonore possono costituire segnali utili per la vita di tutti i giorni. Le frequenze più basse dello spettro acustico includono importanti informazioni che riguardano la presenza e la natura dei suoni provenienti dall'ambiente, come lo squillo del telefono o il rumore delle automobili, e l'intensità e la cadenza della voce, sia la voce del soggetto stesso sia la voce delle persone che il soggetto tenta di capire osservando il movimento delle labbra. La maggior parte dei pazienti con sordità sia congenita sia acquisita in età adulta è rimasta entusiasta anche dei modelli più semplici di protesi a canale singolo.

Si stanno oggi compiendo altri sforzi per migliorare la qualità del suono tramite una elaborazione preliminare più sofisticata e una maggiore fedeltà nel sistema di trasmissione dei segnali dall'unità di regolazione del paziente agli elettrodi impiantati. Un importante progresso è stato realizzato con l'intuizione che molte, se non tutte, le possibilità dei sistemi a canale singolo possono essere realizzate con una stimolazione extracocleare a livello dell'orecchio medio. In questo periodo di incertezza che riguarda i progetti dei sistemi a canali multipli in continua evoluzione e contraddizione, questa scoperta risulta promettente per l'impiego clinico conservativo, ma tuttavia importante, di questa tecnologia in numerosi pazienti (in particolare bambini). È parere comune, tuttavia, che gli stimolatori a canale singolo non possono fornire al sistema nervoso informazioni sufficienti a permettere al paziente di conversare senza l'aiuto della vista.

La diffusione dell'eccitazione da un dato punto di stimolazione nella coclea a parti lontane del nervo acustico è il fattore critico che limita il numero di canali in una protesi cocleare a canali multipli. Il problema è particolarmente difficile nel caso di contatti monopolari *(a sinistra)*. L'aumento dell'intensità della corrente di stimolazione prodotta da un contatto di questo tipo fa sì che la corrente si diffonda all'interno delle concamerazioni piene di liquido dell'orecchio interno, secondo una simmetria radiale, così che l'eccitazione possa raggiungere parti del nervo acustico che normalmente innervano diversi punti della coclea e dia origine in questo modo a differenti sensazioni di tono. Correnti di stimolazione provenienti da altri contatti in queste regioni interferirebbero a loro volta con la diffusione a simmetria radiale e renderebbero difficile la risoluzione dei canali separati di stimolazione oltre a produrre in essi del rumore. Un modo per delimitare la diffusione della corrente di stimolazione è quello di utilizzare contatti bipolari radiali, che tendono a produrre un'eccitazione molto più localizzata *(a destra)*.

L'efficacia delle protesi a canali multipli oggi disponibili è meno chiara. I dispositivi hanno una loro complessità intrinseca, e solo alcuni tra i pazienti che li utilizzano sono stati studiati attentamente. La causa della loro sordità è, inoltre, molto varia così come diverse sono le condizioni del nervo acustico superstite.

Quando un piccolo gruppo di terminazioni del nervo acustico lungo la coclea viene attivato con un dispositivo a canali multipli (con una stimolazione monopolare a bassa intensità o con una stimolazione bipolare), la sensazione di rumore assume un tono definito. Quando vari gruppi fisicamente distinti vengono attivati in sequenza il soggetto incontra scarse difficoltà nell'ordinarli in una scala musicale. È importante rendersi conto che il termine psicofisico «tono» si riferisce a un giudizio soggettivo espresso dall'osservatore, generalmente su una forma d'onda acustica molto complessa che può avere poca o anche nessuna energia alla frequenza corrispondente al tono rilevato. I pazienti con protesi a canali multipli hanno paragonato quello che sentono allo schiamazzare delle anitre o allo sbattere dei bidoni delle immondizie. Questi suoni possono tutti essere classificati in base al tono, ma sono molto differenti dai toni puri percepiti nel caso di onde acustiche sinusoidali.

Non è chiaro se queste sensazioni uditive complesse siano combinabili allo stesso modo come vengono combinate, in esperimenti di simulazione del linguaggio, le onde sinusoidali pure. La parte di sistema nervoso deputata alla funzione uditiva normale ha una straordinaria capacità di ricavare da segnali con rumore l'informazione sottostante e di generalizzare attraverso modelli spettrali solo lontanamente correlati. Per questo motivo si riesce a capire il linguaggio umano parlato sia da un basso sia da un soprano, sia da chi sussurra sia da chi urla, sia infine da soggetti con ampie differenze di accento, di nasalità e di inflessione. La presentazione spettrografica di una data parola è di fatto così complessa e variabile che persino un analista esperto non riesce a identificare la maggior parte delle parole da questo genere di registrazione visiva.

Due differenti indirizzi per l'impiego della dimensione tonotopica dell'informazione acustica sono oggi oggetto di studio in campo clinico. Una è connessa alla simulazione dei suoni vocali tramite canali di frequenza, o *vocoder,* in quanto filtra il segnale acustico suddividendolo in un insieme di bande, ognuna delle quali corrisponde approssimativamente alla percezione del tono, generata da una delle coppie di elettrodi bipolari presenti. Per otto di questi elettrodi vi dovrebbero essere otto attivazioni indipendenti, variabili nel tempo, di gruppi di fibre del nervo acustico e ciascuna attivazione dovrebbe convogliare l'informazione riguardante l'intensità relativa istantanea dell'energia sonora della banda che la riguarda. In teoria la forma d'onda particolare, modulata in ampiezza e usata per attivare ciascun canale, non dovrebbe creare problemi. Per molte ragioni, tuttavia, le onde trasmesse da ciascun filtro passa-banda nell'elaboratore del suono vengono di solito applicate direttamente agli elettrodi. Si eseguono inoltre fasi aggiuntive di pre-elaborazione per compensare lo stretto intervallo dinamico e la sensibilità dei neuroni, dipendente dalla frequenza, alle correnti elettriche.

Si presume che il numero di canali di stimolazione indipendenti e paralleli avrà un effetto significativo sull'intelligibilità del linguaggio. Questo effetto sarà evidente, tuttavia, solo se ciascun canale attiverà realmente un gruppo di fibre del nervo acustico a localizzazione spaziale ben precisa entro gli intervalli dinamici e gli schemi temporali che si incontrano nell'udito normale.

Almeno cinque diversi gruppi di ricercatori hanno studiato pazienti con protesi contenenti da tre a 12 canali di stimolazione ciascuno. In generale, le differenze che esistono riguardo all'intervento chirurgico, nella forma degli elettrodi e nella selezione della forma d'onda, unite alla variabilità individuale dei pazienti, impediscono qualsiasi con-

fronto significativo. In alcuni casi un singolo paziente è stato esaminato variando sistematicamente la suddivisione dello spettro del linguaggio in bande e canali diversi. Canali addizionali conducono a netti e immediati miglioramenti nel riconoscimento delle parole da parte di soggetti portatori di elettrodi che permettono una simile attivazione localizzata. Con una stimolazione su appena quattro canali è stata osservata un'identificazione di parole da un campione casuale pari all'80 per cento del totale, un livello che si avvicina alla riabilitazione funzionale, tenuto conto della ridondanza e degli spunti forniti dal contesto in una conversazione ordinaria.

Una soluzione alternativa alla codificazione tonotopica impiegata nella maggior parte dei dispositivi a canali multipli si basa su una schiera di elettrodi a più canali che varia in modo continuo il punto di applicazione di un singolo canale di stimolazione. In questo schema sia gli elettrodi sia il sistema di trasmissione sono molto più semplici, mentre l'elaborazione del segnale vocale è molto più complessa. Anche elettrodi monopolari molto ravvicinati tendono a dare una sequenza ordinata di sensazioni tonotopiche, quando sono attivati singolarmente, evidentemente perché il sistema nervoso centrale ha scarse difficoltà a identificare il punto centrale in un gradiente relativamente ampio di neuroni attivati.

Le modulazioni di ampiezza e di frequenza di certe componenti dello spettro del linguaggio, come le formanti determinate dalla forma della cavità orale, contengono informazioni particolarmente utili per la discriminazione di molti elementi acustici del linguaggio. Un elaboratore del linguaggio che utilizza un chip di microcalcolatore può identificare e seguire questa caratteristica spettrale, nota come seconda formante, e selezionare la migliore posizione e intensità di stimolazione, basandosi su una mappa memorizzata delle sensazioni prodotte da ogni elettrodo in un particolare soggetto. La frequenza della stimolazione applicata al momento al punto attivato può servire a codificare la frequenza fondamentale della vibrazione delle corde vocali, che corrisponde all'intervallo di basse frequenze per il quale i soggetti riferiscono di avvertire una sensazione del tono simile a una vibrazione. In una ventina di pazienti è stato impiantato un sistema di questo tipo, con 22 elettrodi singoli, progettato da Graeme M. Clark e collaboratori all'Università di Melbourne. Anche in questo caso la variabilità tra i pazienti impedisce ogni tentativo di confrontare i risultati ottenuti con elettrodi a canale singolo o a canali multipli; comunque almeno qualche soggetto è riuscito a raggiungere punteggi significativi nei test di riconoscimento delle parole.

Per migliorare la prestazione saranno necessari canali di informazione corrispondenti alla prima e alla terza formante delle vocali e alle consonanti ad alta frequenza. Forse potrebbero essere emessi due o anche tre stimoli simultanei senza una significativa interazione del canale, a patto che possano essere tenuti sufficientemente distanziati. Rimane da vedere se questo modo generale di affrontare il fondamentale alternarsi del problema del numero di canali e di quello della loro indipendenza potrà fornire l'informazione necessaria in una forma che sia compatibile con le capacità di elaborare informazioni da parte di quelle aree del sistema nervoso deputate alla percezione uditiva.

La complessità delle sensazioni evocate da una stimolazione della coclea la più localizzata possibile ha sorpreso e incuriosito molti ricercatori. La teoria della organizzazione tonotopica della coclea nella sua forma più semplice prevede che un'attivazione localizzata determini una sensazione tonale pura, corrispondente alla frequenza di risonanza della membrana basilare in quel punto. Modificazioni della frequenza e della forma d'onda dello stimolo elettrico dovrebbero dar luogo soltanto a semplici modificazioni di intensità e soltanto nella misura in cui tali modificazioni influiscono sulla velocità media di scarica degli impulsi da parte dei neuroni attivi. In che modo, allora, si originano quelle sensazioni dello spettro complesso che sono, ad esempio, i ronzii o i suoni metallici? È possibile controllarle sistematicamente in modo da fornire un'altra forma di trasferimento dell'informazione mediante protesi?

Da tempo è noto che l'attività del nervo acustico trasmessa al cervello contiene informazioni temporali piuttosto particolareggiate sull'esatta fase del movimento della membrana basilare. Per frequenze sonore al di sotto dei 5000 hertz l'esatta collocazione nel tempo di ogni impulso nervoso in una fibra del nervo acustico è strettamente connessa alla fase del movimento meccanico rilevato dalla particolare cellula ciliata che gli fornisce l'unico input. Anche se le fibre devono fare una pausa di due o tre millisecondi tra un impulso e l'altro, un'analisi spettrale dell'attività in ogni singola fibra rivelerà la frequenza della stimolazione che ha attivato quella singola fibra, sia nel caso che la frequenza corrisponda alla frequenza di risonanza caratteristica di quel punto della membrana basilare sia nel caso contrario. Alcuni neuroni cerebrali che ricevono un input dal nervo acustico presentano una struttura sinaptica specializzata, che può trasmettere e conservare queste informazioni temporali ad alta risoluzione nonostante la limitazione nella frequenza globale di scarica degli impulsi, che è di circa 300 impulsi al secondo. Inoltre, sembra che soggetti sottoposti ad analisi psicofisiche siano in grado di estrapolare informazioni spettrali precise anche quando l'intensità dello stimolo acustico causa una saturazione nella velocità di scarica degli impulsi in aree estese della coclea. Negli anni recenti Murray B. Sachs e Eric D. Young della School of Medicine della Johns Hopkins University hanno sottolineato l'importanza potenziale di questa informazione temporale nella discriminazione dei suoni delle vocali. I loro dati indicano che il sistema nervoso potrebbe benissimo essere in grado di estrapolare questa informazione spettrale strettamente congiunta alla fase e che l'incapacità di riprodurre correttamente questi modelli temporali può dare origine a sensazioni di rumore complesse.

Quali sono le informazioni temporali critiche? Se si combinano assieme molti segnali nervosi in rapporto di fase, ciascuno con una bassa velocità globale di scarica, è possibile ricostruire la frequenza del segnale della sorgente acustica. Questo modo di affrontare il problema dà per scontato come un qualsiasi particolare neurone cerebrale, che riceve lo stimolo ad alta frequenza, riconosca quale sia questa frequenza. Un modo di pensare a un neurone ricevente sintonizzato selettivamente su una particolare frequenza di ripetizione del suo tipo di input è quello di immaginarlo progettato in modo che possa operare con una tecnica detta autocorrelazione temporale. Con questo sistema il segnale in arrivo risulta ritardato di un tempo pari a un ciclo della frequenza da riconoscere; i segnali ritardati e no vengono elaborati insieme. Alcuni modelli di percezione uditiva spesso postulano una schiera di questi rilevatori di periodicità nelle prime sinapsi del tronco cerebrale.

Le teorie biofisiche prevedono, e gli esperimenti sugli animali hanno dimostrato, che una stimolazione elettrica con onde di corrente alternata sinusoidali o a impulsi provoca significativi agganci di fase delle fibre del nervo acustico per frequenze fino ad almeno 3000 hertz. I pazienti con protesi cocleari registrano però costantemente modestissimi cambiamenti di tono e solo lievi modificazioni nella qualità del suono allorché la frequenza di stimolazione aumenta al di sopra dei 300 hertz. Anche quando la naturale frequenza di risonanza del punto dove avviene la stimolazione e la frequenza di ripetizione dell'onda dello stimolo coincidono, non si osserva alcun improvviso miglioramento nel debole tono della sensazione acustica.

Queste osservazioni spingono a riesaminare la classe di meccanismi per mezzo della quale le reti neuronali potrebbero derivare informazioni temporali su una scala del tempo molto più precisa rispetto ai loro abituali processi sinaptici o di conduzione. Un modello possibile, proposto da Mark W. White, Michael M. Merzenich e da me, richiede una correlazione spaziale incrociata per scoprire particolari configurazioni del movimento della membrana basilare, causate dal progredire delle onde viaggianti con una data frequenza. Il processo è analogo alla localizzazione dei suoni nello spazio, basata sul ritardo dei differenti segnali ricevuti dall'uno e dall'altro orecchio. I due processi, in effetti, potrebbero avere in comune una parte dello stesso insieme di circuiti nervosi. La risoluzione temporale sia della localizzazione del suono sia della discriminazione delle frequenze in rapporto di fase è dell'ordine dei 10 microsecondi.

Ovviamente la possibilità di controllare o no con una protesi questi processi e le sensazioni a cui danno origine dipenderà dagli input critici per ciascun processo. Essa può esigere per la stimolazione elettrica un controllo spaziale e temporale più preciso di quanto sia possibile sperare. In alternativa, potrebbe risultare che alcuni semplici accorgimenti nella configurazione degli elettrodi, o nella localizzazione temporale dello stimolo, non ancora sperimentati, siano in grado di svolgere il compito.

Nel 1982, in Italia, un gruppo di otologi, che aveva già iniziato separatamente un rapporto di ricerca con l'House Ear Institute di Los Angeles, ha costituito il gruppo ICI (Impianti cocleari Italia) allo scopo di attuare un programma comune per l'applicazione clinica dell'impianto cocleare 3M nel nostro paese, del quale si è già parlato sul fascicolo di luglio 1984 della rivista (p. 17).

Da allora la situazione non è sostanzialmente mutata per quanto riguarda l'Italia. Attualmente sono 20 i pazienti che si sono sottoposti con successo all'impianto cocleare 3M/House. L'età di questi pazienti varia dai 21 ai 55 anni e la loro sordità era insorta, a secondo dei casi, per meningite, virosi, labirintite, otosclerosi, farmaci ototossici, trauma chirurgico.

I successi ottenuti e il grado di esperienza raggiunta permettono oggi di affrontare il problema della sordità anche nei bambini.

I centri presso i quali è possibile sottoporsi all'intervento e alla successiva riabilitazione sono a Bari, Bergamo, Parma, Trento. Nei prossimi mesi è prevista l'apertura di nuovi centri in altre importanti città italiane che si avvarranno dell'esperienza del gruppo ICI e dell'assistenza tecnica della 3M Italia.

Radar a schiera in fase

Per orientare il fascio, questi radar sfruttano l'interferenza tra i segnali emessi da una moltitudine di piccoli elementi di antenna e quindi localizzano e inseguono oggetti senza spostarsi meccanicamente

di Eli Brookner

L'immagine del riflettore parabolico di un radar che esplora senza posa l'orizzonte con il suo fascio di microonde alla ricerca di oggetti lontani è molto frequente nelle pellicole cinematografiche e, quando si tratta di un radar aeroportuale, è comune nella vita di ogni giorno. Eppure, in molti degli impieghi più comuni del radar, come in aviazione, nella difesa e nello spionaggio, il riflettore parabolico guidato meccanicamente sta per lasciare il posto a un nuovo tipo di apparecchiatura. Una schiera piana di piccole antenne identiche, ciascuna in grado di trasmettere e di ricevere segnali, prende il posto del riflettore concavo, e anche se il suo fascio esplora vaste zone di cielo, il radar in se stesso non si muove. Il segnale viene infatti deflesso elettronicamente da un bersaglio all'altro, guidato sulla base del principio di interferenza delle onde. Si tratta della nuova tecnologia della schiera in fase.

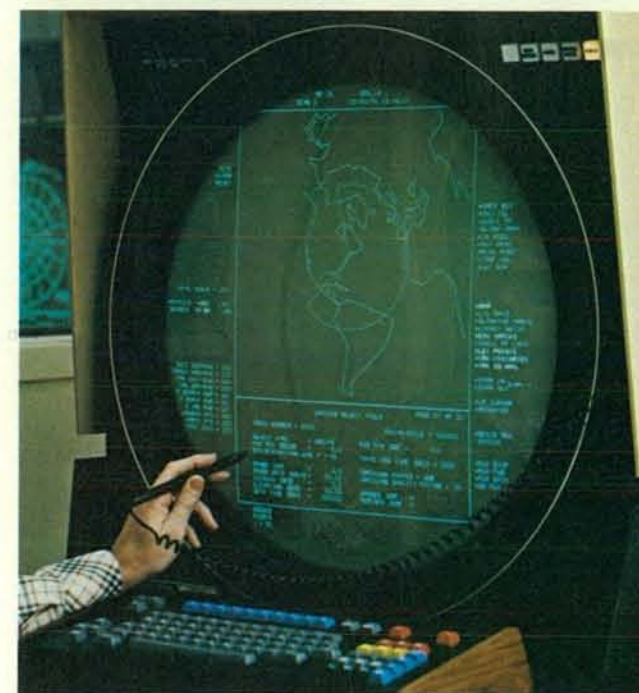
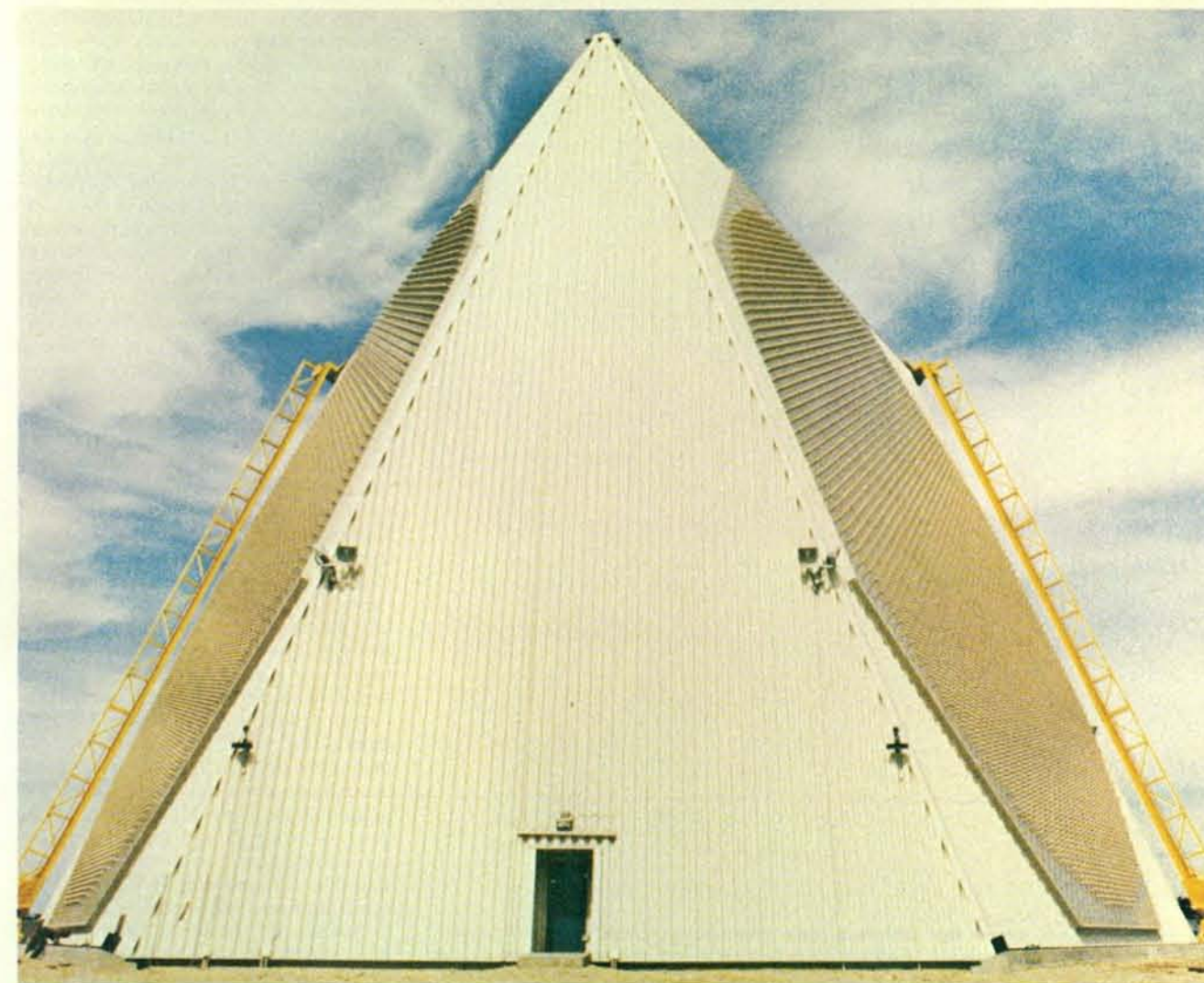
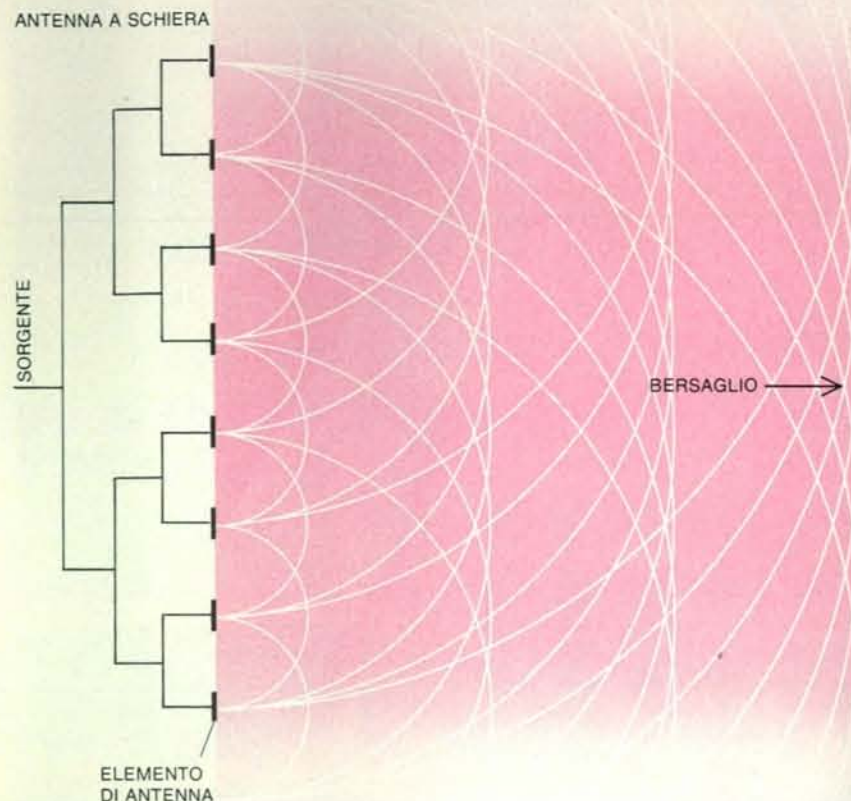
In questa veste i principi basilari del radar restano gli stessi. Tutti i radar funzionano emettendo fasci direzionali di energia radio. Molto spesso l'energia cade nella regione delle microonde compresa tra 300 milioni a 10 miliardi di hertz, anche se alcuni radar per grandissime distanze operano nelle bande di alta frequenza (HF) e di altissima frequenza (VHF), comprendenti rispettivamente frequenze da tre milioni a 30 milioni di hertz e da 30 milioni a 300 milioni di hertz. A seconda della forma dell'antenna, la radiazione forma un fascio sottile, un pennello di raggi, adatto a inseguire con precisione un bersaglio, o un fascio allargato a ventaglio, più adatto a controllare ampie zone di cielo.

Quando il fascio colpisce un oggetto, viene riflesso. Ammesso che l'energia dell'impulso trasmesso, la sensibilità dell'antenna e le caratteristiche di riflettività dell'oggetto siano adeguate, all'antenna ritorna un'eco rivelabile. A seconda del tipo di radar e del tipo di impulso trasmesso, l'eco fornisce diversi tipi di informazione.

La direzione di provenienza dell'eco fornisce le coordinate angolari dell'oggetto e, se il radar trasmette impulsi di energia anziché un segnale continuo, l'intervallo tra la trasmissione di un impulso e l'arrivo della sua eco esprime la distanza dell'oggetto. Alcuni radar sono progettati per misurare anche lo spostamento Doppler dell'eco, cioè la variazione di frequenza del segnale che si ha quando la sorgente (in questo caso il bersaglio) e il ricevitore (la stazione radar) sono in movimento l'una rispetto all'altro. Dallo spostamento Doppler questi radar ricavano la velocità di avvicinamento o di allontanamento di un oggetto rispetto all'antenna.

Per una data distanza, l'intensità dell'eco fornisce alcune indicazioni sulle dimensioni dell'oggetto. Il termine «indicazione» è usato a proposito; due oggetti delle stesse dimensioni, ma di forma differente o fatti di materiali diversi, rimandano echi di intensità nettamente diversa. Per ottenere informazioni più precise sulle dimensioni, alcuni radar trasmettono impulsi così brevi da risultare spazialmente più corti dei bersagli che si intendono colpire, o «illuminare». Se un radar emette energia solo per alcuni miliardesimi di secondo, il fronte dell'impulso avrà percorso solo poche decine di centimetri prima che la sua emissione sia terminata. Un tale impulso può essere più breve, per esempio, di un aereo. I segnali radar vengono riflessi sia dalla parte anteriore, sia da quella posteriore del bersaglio; nel caso di un impulso estremamente breve, le due riflessioni forniranno due echi distinti. L'intervallo tra gli echi corrisponde alla lunghezza del bersaglio.

Un radar convenzionale che esplora ampie zone di cielo può raccogliere informazioni su un gran numero di oggetti. È però inevitabile che intercorra un tempo notevole tra due osservazioni successive dello stesso bersaglio. La frequenza di ripetizione, cioè la frequenza con la quale il radar aggiorna le indicazioni su un dato bersaglio, per la maggior parte dei radar orientati meccanicamente è inferiore alla frequenza di rotazione del riflettore parabolico attorno al proprio asse. Nei radar per il controllo del traffico aereo, ad esempio, la riga verde radiale che spazzola circolarmente lo schermo del tubo a raggi catodici lasciando impresse le posizioni aggiornate e ulteriori informazioni sull'aereo che si trova entro la portata del radar, ruota con la stessa frequenza con cui ruota materialmente il riflettore parabolico del radar. La frequenza di ripetizione di tali radar è tipicamente di circa una scansione dello schermo ogni sei secondi e anche i più avanzati radar militari raramente raggiungono frequenze maggiori di due scansioni al secondo.

Vi sono però circostanze che richiedono letture più frequenti della posizione e del movimento del bersaglio. Un singolo radar orientato meccanicamente può fornire con continuità dati su uno o più oggetti molto vicini seguendone la traiettoria con la propria rotazione in modo da assecondarne il movimento. Tuttavia, per molti impieghi militari e di spionaggio - per esempio l'inseguimento effettuato da bordo di navi di numerosi missili attaccanti da direzioni diverse, o l'esame minuzioso dei diversi componenti del carico utile di un missile balistico intercontinentale in volo di prova - è necessario osservare contemporaneamente e con continuità un gran numero di bersagli. Sino a poco tempo fa, a questo scopo potevano essere utilizzati soltanto gruppi di radar, ciascuno dei quali controllava una parte dei bersagli. Oggi un radar a schiera in fase può svolgere da solo il lavoro di una tradizionale batteria di riflettori parabolici orientati meccanicamente. Per fare un esempio concreto, un radar a schiera in fase, denominato in codice COBRA DANE, che dal sito di installazione sul Mare di Bering osserva le prove dei missili balistici sovietici, può seguire centinaia di bersagli sparsi in una

Il PAVE PAWS, una coppia di schiere in fase *(in alto)*, è un radar bifronte installato alla Otis Air Force Base di Cape Cod. Costruito dalla Raytheon Company, l'impianto ha lo scopo di fornire il primo allarme in caso di attacco da parte di missili balistici lanciati da sommergibili e di servire da mezzo ausiliario per l'inseguimento di satelliti. Al contrario dei riflettori parabolici orientabili dei radar convenzionali, nessun elemento dell'antenna del PAVE PAWS è mobile. Il suo fascio viene invece orientato elettronicamente sfruttando l'interferenza dei segnali separati emessi dai 1792 elementi radianti di ogni faccia larga circa 31 metri. L'attività combinata delle antenne gemelle separate da 120 gradi e ciascuna in grado di esplorare 120 gradi in azimut, fornisce al radar un campo di visibilità di 240 gradi rappresentato nella visualizzazione sintetizzata al calcolatore *(a sinistra)*. Entro questo angolo può seguire quasi simultaneamente un gran numero di oggetti spostando elettronicamente il fascio da un bersaglio all'altro in alcuni milionesimi di secondo. Entro il suo raggio di azione di 3000 miglia nautiche il sistema PAVE PAWS può rivelare oggetti con sezione frontale di 10 metri quadrati. La penna luminosa serve per richiamare sullo schermo i dati e modificare l'informazione negli archivi di dati.

Gli elementi di antenna di una schiera in fase emettono segnali a microonde distinti. Quando tutti gli elementi irradiano esattamente in concordanza di fase, dando luogo a creste d'onda che si muovono in sincronia in avanti, le onde si sovrappongono lungo l'asse perpendicolare allo specchio della schiera. Le onde interferiscono costruttivamente e danno luogo a un intenso segnale-somma che si manifesta sotto forma di un fascio diretto frontalmente.

zona dello spazio di 120 gradi in azimut e di circa 80 gradi in elevazione. In realtà li osserva simultaneamente, spostando elettronicamente il fascio di radiazione da un bersaglio all'altro in tempi dell'ordine dei microsecondi (milionesimi di secondo).

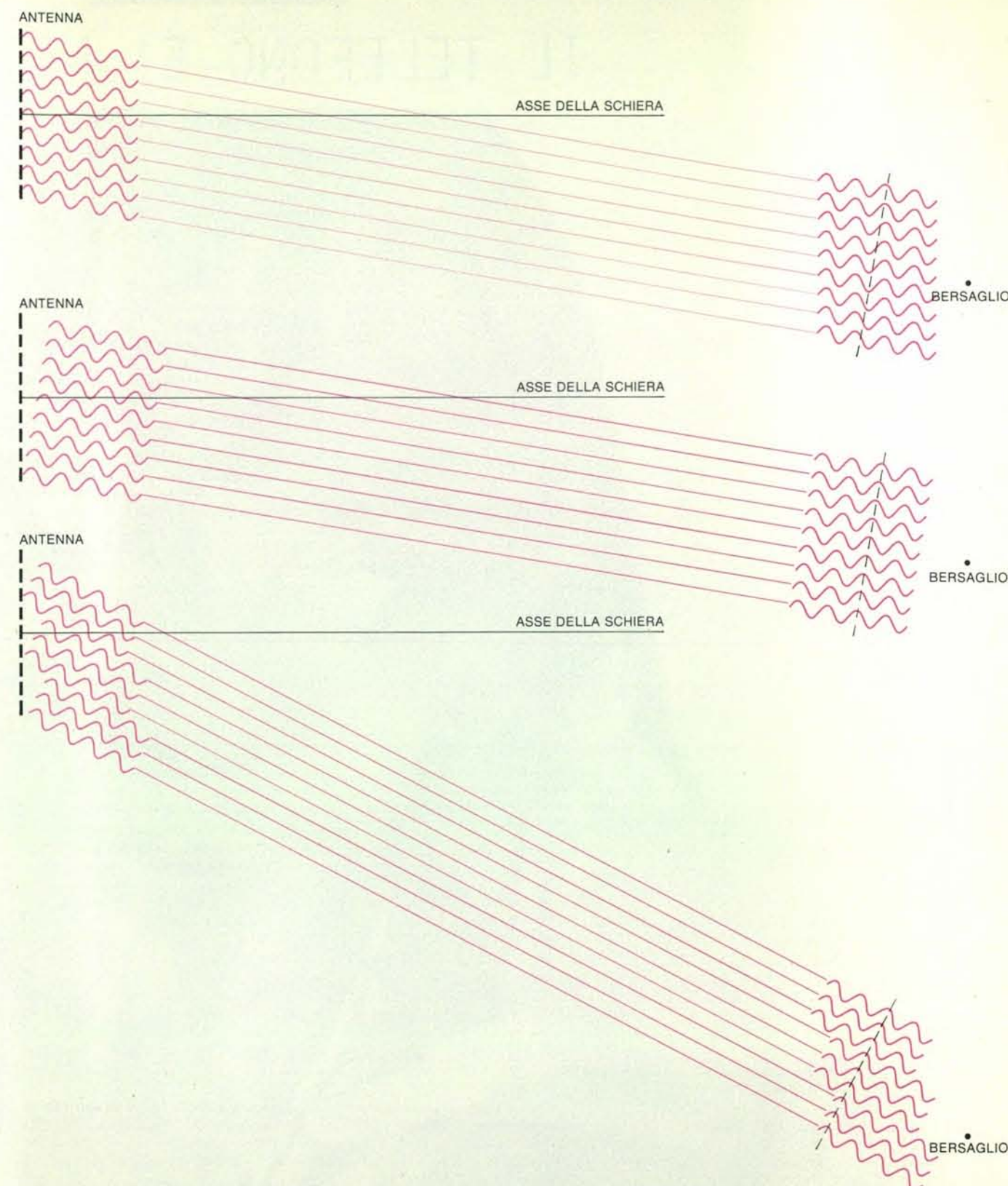
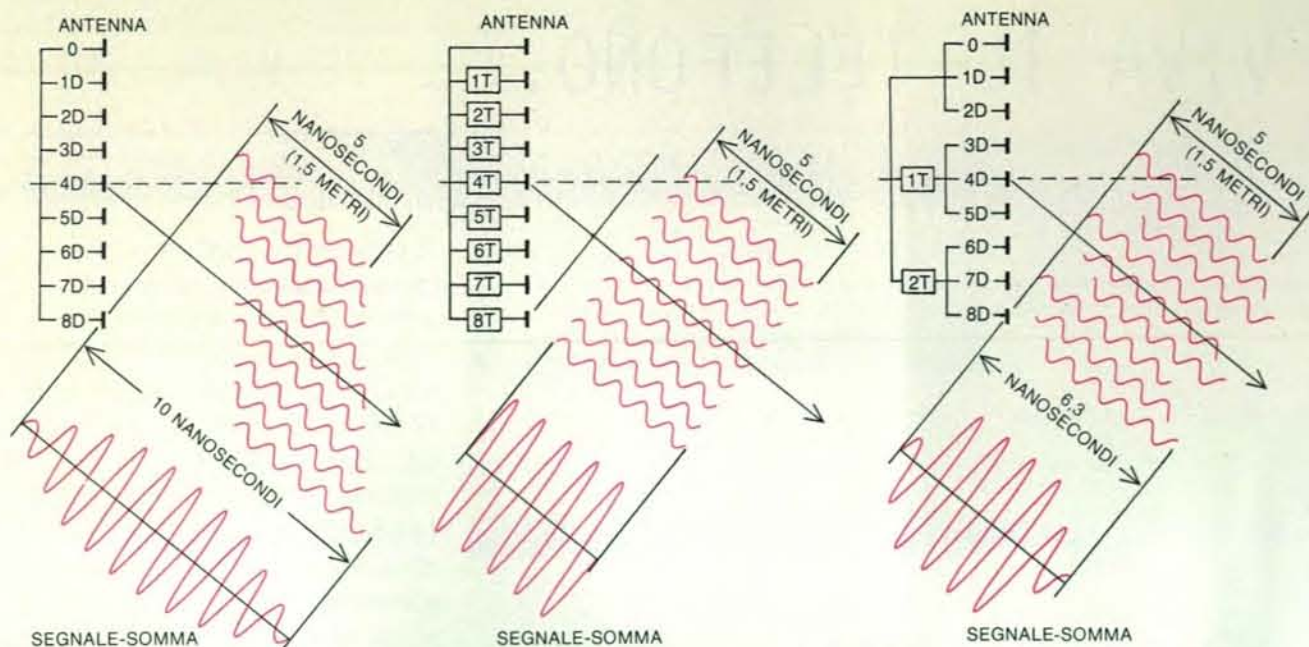
La guida elettronica del fascio che è alla base di queste straordinarie capacità sfrutta un semplice principio fisico. Quando sorgenti adiacenti irradiano simultaneamente energia alla stessa frequenza, le onde risultanti interferiscono. Quando due onde provenienti da sorgenti distinte si sovrappongono, la fase relativa, che è una misura della relazione esistente tra le rispettive forme d'onda, determina in che modo avviene l'interferenza. Se creste e creste coincidono, e così pure i ventri, le onde si combinano in modo da magnificare il segnale risultante, un fenomeno chiamato interferenza costruttiva. Se le forme d'onda sono sfasate e le creste e i ventri non coincidono, ne deriverà un'interferenza distruttiva: i due segnali daranno luogo a un segnale indebolito, o anche si annulleranno completamente tra loro.

Una schiera in fase è costituita da una disposizione piana e regolare di elementi radianti; ogni elemento viene alimentato con un segnale a microonde e tutti i segnali hanno uguale ampiezza e sono coerenti in fase. Il segnale viene generato da un oscillatore centrale e amplificato da transistori o da speciali tubi per microonde, come per esempio tubi a onda viaggiante. Se tutti i segnali lasciano la schiera in fase, si sommano, cioè interferiscono costruttivamente, in tutti i punti di una linea perpendicolare allo specchio, ossia al fronte, della schiera. Di conseguenza il segnale risultante sarà intenso e sarà in grado di produrre echi rivelabili da oggetti che si trovino sul suo percorso, lungo l'asse della schiera ed entro un piccolo cono attorno a esso.

Ad angoli più grandi rispetto all'asse della schiera, i segnali provenienti dai diversi elementi radianti devono percorrere distanze diverse per raggiungere un bersaglio. Ne risulta che le loro fasi relative sono modificate ed essi interferiscono distruttivamente, indebolendo o eliminando il fascio. Pertanto, all'esterno del ristretto cono in cui si ha interferenza costruttiva, gli oggetti non producono echi rivelabili. Per le caratteristiche delle figure di interferenza, l'ampiezza di questo cono è direttamente proporzionale alla lunghezza d'onda di funzionamento e inversamente proporzionale alle dimensioni della schiera. Se tutti gli elementi radianti sono in fase, il fascio viene in realtà orientato frontalmente, nella direzione dell'asse della schiera.

Supponiamo ora che i segnali emessi da ciascun elemento radiante siano ritardati elettronicamente di quantità che aumentino regolarmente lungo lo specchio della schiera. Ogni ritardo fa sì che il segnale corrispondente sia posposto di una frazione di lunghezza d'onda rispetto al segnale dell'elemento adiacente. Ne risulta una variazione delle fasi relative dei segnali. Ora, la zona in cui i singoli segnali si sommano costruttivamente e producono un intenso segnale in grado di rivelare un bersaglio, non si trova più frontalmente lungo l'asse dell'antenna, ma risulta spostato lateralmente nella direzione dell'aumento del ritardo di fase. L'angolazione del fascio dipende dal valore dello spostamento di fase, o sfasamento, dalle dimensioni della schiera e dalla lunghezza d'onda dei segnali. Il fascio assume, anche in questo caso, la forma di uno stretto cono circondato da una zona di interferenza distruttiva. Il fascio radar è stato deflesso senza muovere fisicamente l'antenna.

Quando un'eco viene riflessa da un bersaglio che si trova lungo questo nuovo asse di puntamento creato elettronicamente dai ritardi progressivi di fase, il sistema circuitale che ha ritardato i segnali trasmessi introduce un nuovo insieme di ritardi nei diversi segnali ricevuti dai singoli elementi radianti. Poiché il fronte dell'onda di ritorno spazzola obliquamente lo specchio della schiera, gli elementi che hanno irradiato per ultimi, quelli più vicini al bersaglio, ricevono per primi l'impulso riflesso. Pertanto, lo stesso insieme di ritardi che ha orientato l'impulso trasmesso riporta in fase tutti i segnali che costituiscono l'eco, pronti per l'elaborazione che porta alle informazioni sul bersaglio. Il puntamento basato sul ritardo di fase mette in grado una schiera tipica di spostare il fascio sino a 60 gradi dall'asse perpendicolare, fornendo al radar un campo trasversale di visibilità di 120 gradi: un terzo della linea dell'orizzonte e, se lo specchio della schiera è convenientemente inclinato, dall'orizzonte a ben oltre lo zenit. Poiché il puntamento non richiede un riassetto meccanico, il movimento del fascio lungo tutto il campo visuale richiede solo pochi microsecondi. Con un calcolatore che determina i valori necessari di sfasamento per ottenere l'angolazione del fascio desiderata e che controlla il sistema circuitale di ritardo,

ANTENNA
ASSE DELLA SCHIERA
BERSAGLIO
ANTENNA
ASSE DELLA SCHIERA
BERSAGLIO
ANTENNA
ASSE DELLA SCHIERA
BERSAGLIO

Il puntamento elettronico del fascio richiede spostamenti progressivi della fase del segnale. Quando gli elementi di antenna irradiano in concordanza di fase, ciascuno di essi emette un segnale che si propaga in tutte le direzioni. In direzione diversa dall'asse di puntamento, cioè dall'asse della schiera, le creste e i ventri dei segnali non coincidono *(in alto)*. In questa direzione i segnali sono sfasati; l'impulso radar è perciò indebolito da un'interferenza distruttiva e non produce un'eco rivelabile. Per spostare la regione di interferenza costruttiva verso l'oggetto, cioè per puntare effettivamente il fascio, i segnali vengono ritardati di frazioni di lunghezza d'onda. Ne consegue una serie di sfasamenti successivi che aumentano lungo lo specchio della schiera. I segnali sfasati coincidono ora lungo una direzione che forma un certo angolo con l'asse della schiera *(al centro)*, interferendo costruttivamente e producendo così un intenso segnale. Per deflettere ulteriormente il fascio, l'incremento tra due sfasamenti successivi deve essere maggiore. Per ridurre il numero di circuiti necessari, i ritardi non sono mai maggiori di un'intera lunghezza d'onda, ma si hanno incrementi graduali che vanno da un ritardo zero al massimo sfasamento di circa una lunghezza d'onda e che si ripetono lungo lo specchio della schiera *(in basso)*. Si ottiene così un fascio in cui le creste dei segnali coincidono per un angolo molto grande rispetto all'asse della schiera, producendo interferenza costruttiva e un intenso segnale.

I ritardi di tempo impediscono che un impulso radar molto breve, necessario per determinare le dimensioni di un oggetto, venga allungato e distorto quando viene trasmesso secondo una direzione molto angolata rispetto all'asse perpendicolare alla schiera. Un impulso di cinque nanosecondi (cinque miliardesimi di secondo) misura 1,5 metri dall'inizio alla fine quando viene irradiato frontalmente, ma questo valore può aumentare di molte volte *(a sinistra)* quando viene deflesso a grandi angoli per effetto degli sfasamenti rappresentati da multipli di *D* (per chiarezza non sono stati riportati i ritardi delle singole forme d'onda). I ritardi progressivi, rappresentati da multipli di *T* e ammontanti a molte volte i ritardi usati per la deflessione di fase, consentono di conservare forma e dimensioni dell'impulso *(al centro)*. Per evitare la complessità di circuiti di ritardo di tempo distinti per ciascuno delle centinaia o migliaia di elementi di antenna, molte schiere in fase sono suddivise in sottoschiere, ciascuna delle quali è controllata con lo spostamento di fase. Ad alti angoli di puntamento, le sottoschiere producono segnali distorti, che però sono coordinati attraverso i ritardi di tempo in modo che le distorsioni non si sommino lungo lo specchio della schiera. Le dimensioni dell'impulso restano abbastanza precise e sono necessari solo pochi circuiti di ritardo di tempo *(a destra)*.

una schiera in fase come il COBRA DANE può inseguire contemporaneamente diverse centinaia di bersagli.

La strumentazione elettronica essenziale per il puntamento a ritardo di fase, ossia il dispositivo che crea il giusto ritardo nel segnale a microonde inviato a ciascun elemento, è uno «sfasatore», costituito da due cavi o guide d'onda di dimensioni esattamente definite e funziona nel modo seguente. Aumentando la lunghezza del cavo lungo il quale viaggia il segnale dall'oscillatore o dall'amplificatore all'elemento radiante si introduce un ritardo nella trasmissione del segnale. Non è però possibile far variare all'infinito la lunghezza dei cavi di alimentazione degli elementi radianti di una schiera in fase per produrre ritardi di fase variabili con continuità. Lo sfasamento viene invece realizzato per gradi in modo discontinuo. A ogni elemento radiante è connessa materialmente una gerarchia di cavi formata da elementi di diversa lunghezza. In ciascun circuito viene commutata una particolare combinazione di questi cavi in modo da ottenere gli sfasamenti necessari per produrre la deflessione del fascio voluta.

Il radar COBRA DANE, per esempio, usa combinazioni di tre sfasamenti distinti, ottenuti con tre cavi («*bit*») di lunghezza diversa. Ogni sfasatore contiene quindi tre diverse linee di trasmissione «a striscia», un tipo di guida d'onda che canalizza l'energia a microonde lungo una stretta striscia di rame centrata tra due piani di massa in rame. Una delle linee di trasmissione a striscia aggiunge mezza lunghezza d'onda al percorso del segnale, circa 15 centimetri alla frequenza operativa del COBRA DANE che è di circa un gigahertz (un miliardo di cicli al secondo). Essa introduce quindi uno sfasamento di 180 gradi rispetto al segnale non ritardato. Un'altra linea di trasmissione a striscia introduce nel segnale un ritardo di un quarto di lunghezza d'onda, per uno sfasamento di 90 gradi, e la terza è equivalente a un ottavo di lunghezza d'onda, ovvero a uno sfasamento di 45 gradi. In combinazione opportuna, le tre linee di trasmissione a striscia permettono di modificare la fase del segnale di un valore pari a qualsiasi multiplo intero di 45 gradi, da zero a 315 gradi.

Sembrerebbe che i ritardi di fase per passi discontinui implichino la formazione di punti ciechi, ossia di zone di silenzio. Come possono gli unici otto possibili ritardi di fase, a intervalli di 45 gradi, permettere di modificare con continuità l'orientazione del fascio radar? La risposta sta nelle proprietà caratteristiche delle figure di interferenza. Ogni volta che la differenza di fase tra i segnali irradiati a estremità opposte della schiera viene incrementata di 360 gradi, ovvero di una lunghezza d'onda, la regione di interferenza costruttiva che dà luogo al fascio si sposta di un tratto circa uguale alla sua larghezza. Per spostare il fascio dalla posizione perpendicolare alla schiera, che si ha quando tutti i segnali sono in fase, a una posizione contigua, senza che vi siano punti ciechi, lo sfasamento totale lungo tutto lo specchio della schiera deve essere di circa 360 gradi.

Che gli sfasamenti individuali aumentino con continuità lungo lo specchio della schiera o con incrementi finiti di 45 gradi è irrilevante; incrementi più grossolani producono solo una lieve caduta dell'intensità del fascio e della sensibilità dell'antenna. Per raggiungere una maggiore orientabilità, un'antenna equipaggiata con sfasatori a tre *bit* può basarsi su sfasamenti complessivi minori: per esempio, di 180 gradi divisi in quattro intervalli di 45 gradi.

Se il fascio deve essere deflesso di una quantità maggiore della sua larghezza, lo sfasamento complessivo lungo lo specchio della schiera deve essere maggiore di 360 gradi. Per il carattere periodico delle onde elettromagnetiche, uno sfasamento di multipli di lunghezza d'onda è equivalente a uno sfasamento di 360 gradi. Per una variazione di fase totale superiore a 360 gradi l'incremento lineare del ritardo di fase da zero a 360 gradi deve essere ripetuto più di una volta lungo lo specchio della schiera. La prima serie di ritardi di fase sposta la fase di un valore totale corrispondente a una lunghezza d'onda; la seconda serie la sposta di un'altra lunghezza d'onda, e così via. Un grafico del ritardo di fase in funzione della distanza lungo lo specchio della schiera avrebbe l'aspetto di una serie di denti di sega; quanto più i loro lati inclinati sono ripidi e quanto maggiore è il loro numero, tanto più netta sarà la deflessione del fascio.

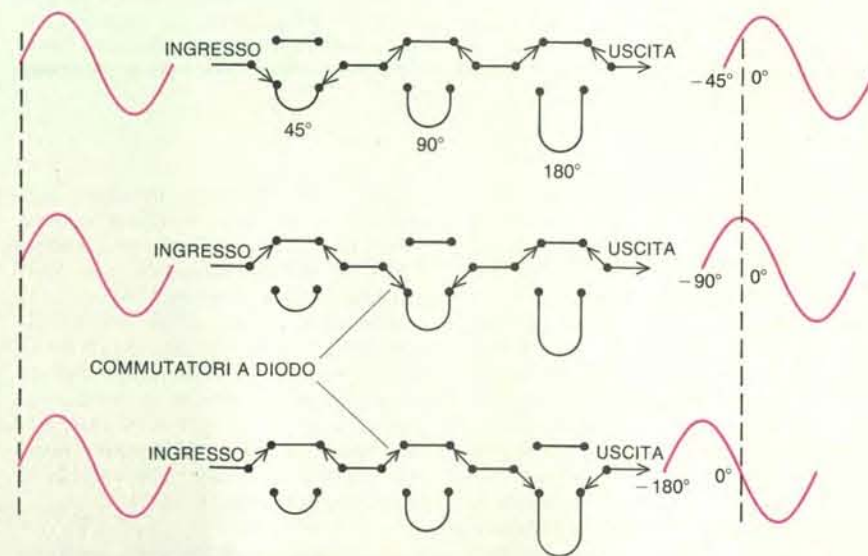
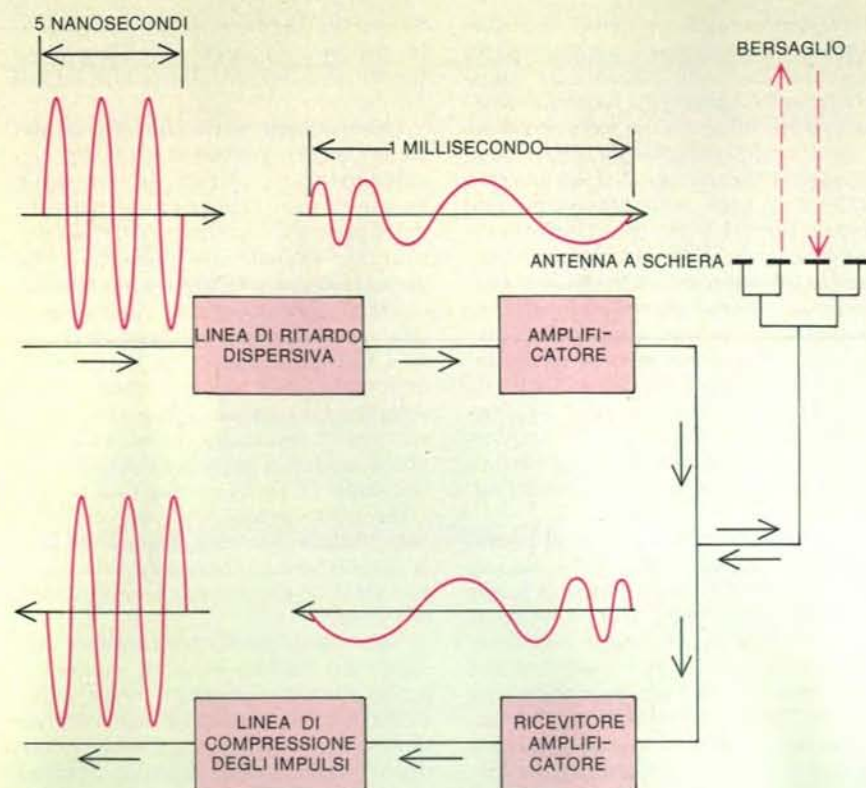
Semplici considerazioni geometriche mostrano che l'area equivalente dell'antenna diminuisce all'aumentare dell'angolo di orientazione. Ne risulta che la sensibilità di una schiera in fase nei confronti degli echi di ritorno da un bersaglio cade rapidamente per angoli maggiori di 60 gradi rispetto all'asse della schiera. Di per se stessa, una schiera in fase non può uguagliare la capacità di esplorazione omnidirezionale di un riflettore parabolico fatto ruotare meccanicamente. Una soluzione consiste nel collegare più schiere rivolte in varie direzioni. Un altro modo di aumentare la copertura di una schiera in fase è quello di montare la schiera orizzontalmente sotto una lente a cupola che rifranga il fascio; una sistemazione che ne aumenta la deflessione angolare. Quando la schiera produce un fascio a 60 gradi rispetto allo zenit, la lente potrebbe inclinarlo ulteriormente fino a farlo viaggiare ad angolo retto, esattamente in direzione dell'orizzonte. La lente fornisce quindi alla schiera la capacità di esplorare l'intero emisfero celeste. Una lente di questo tipo è fatta di materiale ceramico o plastico speciale che rifrange le microonde; può anche assumere la forma di un secondo insieme di sfasatori, predisposti per introdurre un ritardo di fase addizionale sul segnale ricevuto dalla schiera.

Quando si ricorre al ritardo di fase per dirigere un breve impulso di energia radar a un grande angolo rispetto all'asse della schiera, l'impulso uscente è inevitabilmente distorto: allungato nello spazio e dilatato nel tempo. Supponiamo che un'antenna emetta un impulso che duri cinque nanosecondi (cinque miliardesimi di secondo). Se il radar è puntato frontalmente, l'impulso che lascia l'antenna ha una sezione trasversale rettangolare, di larghezza uguale a quella della schiera e di lunghezza uguale alla distanza percorsa dalle onde elettromagnetiche in cinque nanosecondi, cioè 1,5 metri. Se, d'altra parte, lo sfasamento ha prodotto una netta deflessione del fascio rispetto all'asse della schiera, l'impulso assumerà l'aspetto di un parallelogramma. Dal punto di vista di un bersaglio la lunghezza dell'impulso sarà maggiore di 1,5 metri poiché i segnali che lo compongono, invece di colpirlo simultaneamente, lo raggiungeranno in successione. Anche l'eco di ritorno sarà allungata.

Generalmente, i radar, per localizzare e inseguire bersagli, trasmettono impulsi molto più lunghi - per esempio lunghi 1000 nanosecondi - e quindi la distorsione di alcuni nanosecondi è un effetto trascurabile. L'allungamento ha scarso effetto sulla capacità del radar di ottenere dall'eco informazioni sulla posizione e sulla velocità del bersaglio. Per risolvere oggetti che viaggiano in formazione compatta è però necessario trasmettere impulsi brevi. Un impulso di breve durata è essenziale anche per i radar progettati per misurare le dimensioni di un bersaglio in base agli echi distinti che ritornano rispettivamente dalla parte anteriore e dalla parte posteriore del bersaglio quando viene illuminato con un impulso estremamente breve. Se l'impulso trasmesso viene troppo allungato, gli echi che dovrebbero essere distinti possono fondersi, oscurando l'informazione.

Un procedimento analogo al puntamento basato sul ritardo di fase supera questa difficoltà conservando la forma originale dell'impulso. Nel puntamento a ritardo di fase è necessario introdurre nei segnali ritardi pari a frazioni di una lunghezza d'onda; i ritardi necessari per evitare l'allungamento dell'impulso corrispondono a multipli di una lunghezza d'onda. In pratica, ai singoli segnali vengono dati dei «vantaggi» proporzionali alla loro distanza dal bersaglio; il risultato è lo stesso che si avrebbe se si facesse ruotare la schiera in modo da portare il bersaglio sull'asse di puntamento. Questo procedimento di puntamento a ritardo di tempo, come quello basato sull'aumento del ritardo di fase di un segnale, produce un fascio con l'angolo di puntamento voluto, coerente e quindi intenso.

Questi lunghi ritardi, che equivalgono a qualche metro di percorso del segnale, richiedono che cavi di proporzionale lunghezza siano incorporati nel percorso del segnale dall'oscillatore o dall'amplificatore all'elemento radiante. Una grande schiera in fase può comprendere molte migliaia di elementi radianti; se a ciascuno di essi fosse collegato un circuito a ritardo di tempo, ne risulterebbe un'installazione radar enormemente complessa e costosa. I progettisti di sistemi radar raggiungono invece un compromesso tra la precisione della forma dell'impulso, anche per angoli di deflessione molto grandi, e la semplicità di realizzazione. Si ottiene una schiera nella quale il puntamento del fascio è regolato sia dallo sfasamento, sia dal ritardo di tempo.

Nel COBRA DANE, per esempio, ciascuno dei 15 360 elementi radianti è connesso a uno sfasatore a tre *bit*; quindi, ciascuno dei segnali subisce il ritardo di fase individualmente. Nella ricerca del bersaglio, il radar trasmette impulsi di 1000 nanosecondi e viene guidato solo attraverso il ritardo di fase. Poiché, però, lo scopo del radar è quello di controllare le prove dei missili balistici sovietici, esso deve raccogliere informazioni sulle dimensioni degli oggetti una volta che siano stati localizzati. A questo

Gli sfasatori incanalano i segnali a microonde che alimentano ciascun elemento di antenna attraverso cavi di lunghezza variabile. I cavi ritardano il segnale, spostando così la fase relativa dell'uscita. Poiché le microonde viaggiano a velocità costante, il ritardo prodotto da ciascun percorso è proporzionale alla sua lunghezza. L'illustrazione mostra i tre ritardi fondamentali che ciascun sfasatore è in grado di introdurre. Il più corto dei tre percorsi corrisponde a un ritardo di 45 gradi, ovvero a un ottavo di lunghezza d'onda *(in alto)*. Se questo percorso viene tagliato fuori e il successivo viene inserito, lo spostamento di fase è di 90 gradi, ovvero un quarto di lunghezza d'onda *(al centro)*. Il più lungo dei tre percorsi produce un ritardo di 180 gradi, ovvero di mezza lunghezza d'onda *(in basso)*. Diverse combinazioni dei tre percorsi possono produrre spostamenti di fase, a intervalli di 45 gradi, sino a 315 gradi. Per ciascun angolo di puntamento del fascio radar il calcolatore centrale calcola il ritardo di fase da assegnare a ciascuno degli elementi radianti e commuta i percorsi degli sfasatori nella giusta combinazione.

La codificazione e la compressione degli impulsi mettono in grado un radar di trasmettere impulsi lunghi, ottenendo informazioni sulle dimensioni e sulla distanza del bersaglio altrettanto precise di quelle ottenute con un impulso breve. Questo richiede generalmente per un raggio adeguato una potenza di trasmissione impraticabilmente elevata. Nella codificazione degli impulsi, un segnale molto breve costituito da uno spettro di frequenze passa attraverso una linea di ritardo dispersiva che ne ritarda le componenti di quantità proporzionali alla frequenza. Nel processo l'impulso viene allungato; per esempio, un impulso di cinque nanosecondi potrebbe essere allungato di 200 000 volte raggiungendo la durata di un millisecondo. Poi, viene amplificato e trasmesso; dal bersaglio ritorna un'eco e anch'essa viene amplificata. Una linea di compressione degli impulsi ritarda l'eco di quantità inversamente proporzionali alla frequenza delle componenti in modo da ridurre il segnale alla sua lunghezza originale di cinque nanosecondi. L'eco così compressa fornisce tutte le informazioni che sarebbero state disponibili se fosse stato trasmesso un impulso inalterato di cinque nanosecondi.

scopo la sua antenna è organizzata in 96 sottoschiere, o blocchi, ciascuna di 160 elementi radianti. Non appena il radar ha localizzato un bersaglio, comincia a trasmettere impulsi estremamente brevi mentre nei segnali inviati a ciascuna sottoschiera vengono introdotti ritardi di tempo. Il sistema circuitale per il ritardo di tempo è simile, ma di scala maggiore, a quello degli sfasatori: consiste di cavi coassiali di lunghezze determinate e ogni combinazione di questi cavi può essere connessa al circuito in modo da fornire un ritardo di tempo corrispondente a una variazione di percorso compresa tra una lunghezza d'onda e 64 lunghezze d'onda, circa 20 metri alla frequenza operativa, circa un gigahertz, del COBRA DANE.

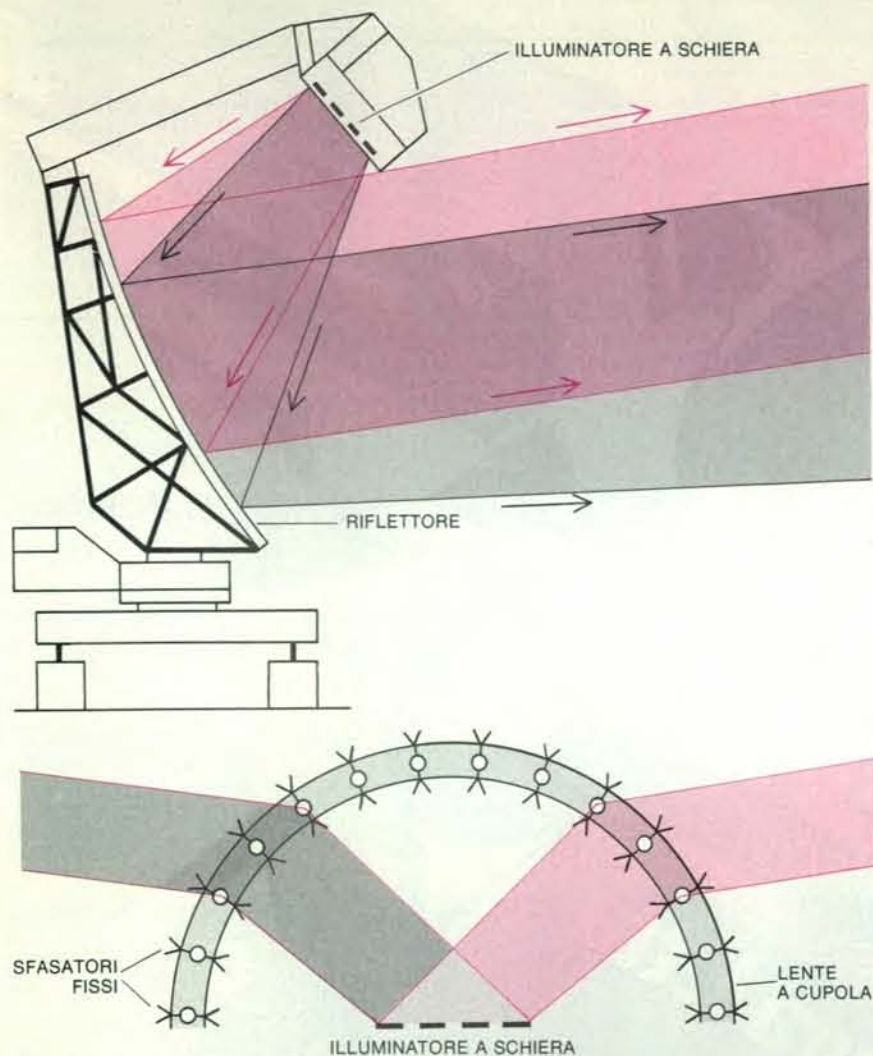
Poiché le sottoschiere hanno un diametro di soli 2,7 metri rispetto ai 28,5 metri della schiera completa, la distorsione dovuta a ciascuna sottoschiera per grandi angoli di puntamento è moderata. Ciascuna sottoschiera produce un segnale a forma di parallelogramma, ma questi segnali si compongono dopo essere stati sottoposti a un ritardo di tempo e quindi le singole distorsioni non si sommano. Ne risulta una ragionevole conservazione della forma dell'impulso e la possibilità di ridurre il numero dei dispositivi di ritardo di tempo da 15 360 a soli 96. In termini di materiali, il sistema di puntamento a ritardo di tempo del COBRA DANE richiede circa un chilometro e mezzo di cavo addizionale e non i 160 chilometri di cavo che sarebbero altrimenti necessari.

La sostituzione di un'unica antenna mobile con una schiera fissa di elementi radianti può offrire, oltre a quello del puntamento elettronico, ulteriori vantaggi di interesse pratico. Uno di questi è l'affidabilità. Il funzionamento di una schiera fissa non è influenzato dai guasti di componenti meccanici come ingranaggi e motori; inoltre la maggior parte dei radar orientati meccanicamente impiega per l'amplificazione dei segnali uno o al più solo alcuni tubi per microonde molto grandi. Per fare un esempio, il radar Martello della Marconi, un radar per la difesa aerea di fabbricazione britannica, ha come elemento essenziale un unico tubo con potenza di uscita di circa tre megawatt. Se il tubo si guasta, anche l'intero sistema cessa di funzionare. Questi radar usati nella difesa aerea e nello spionaggio possono però essere rapidamente commutati a fonti ausiliarie di energia.

Il COBRA DANE, invece, è alimentato da 96 tubi, ciascuno dei quali fornisce 160 chilowatt di potenza. Il segnale in uscita da ciascun tubo passa attraverso un divisore che lo ripartisce tra i 160 elementi radianti di una sottoschiera. Il guasto di un tubo del COBRA DANE blocca pertanto solo un novantaseiesimo della schiera, lasciando il radar nel suo complesso ancora funzionante, con solo una piccola riduzione di prestazioni. Inoltre, i tubi più piccoli sono sostituibili più facilmente dell'unico tubo grande del radar Martello.

Le schiere in fase a stato solido possono raggiungere livelli di affidabilità e di praticità di manutenzione ancora superiori. I radar denominati in codice PAVE PAWS, progettati per rivelare, da posizioni favorevoli sul Cape Cod e in California, e da siti strategici in Georgia e nel Texas, missili balistici lanciati dal mare, sono alimentati da transistori. In moduli distinti per ogni elemento radiante sono collegati in parallelo quattro transistori da 100 watt. Perciò, su ciascuna delle due facce gemelle dell'antenna PAVE PAWS l'amplificazione del segnale è distribuita tra 1792 elementi d'antenna e non tra 96 tubi, cosicché le conseguenze di un singolo guasto sulle prestazioni del radar sono ancora minori. Inoltre, il tempo medio tra due guasti successivi di un modulo a stato solido è notevolmente maggiore di quello di un tubo del COBRA DANE: 100 000 ore rispetto a 20 000 ore. Quando i moduli, lunghi 30 centimetri e alimentati a 28 volt, si guastano, sono ancora più facili da sostituire dei tubi alti 1,5 metri e alimentati a 40 000 volt del COBRA DANE.

Nel PAVE PAWS, come in molti radar a stato solido, l'amplificazione ha luogo dopo che il segnale è stato suddiviso tra gli elementi di antenna e dopo aver subito lo sfasamento. Ne segue che in questo caso non hanno luogo le perdite di potenza che si hanno quando un segnale amplificato passa per un circuito divisore e per uno sfasatore. Tuttavia, per questo guadagno di efficienza e per tutti gli altri vantaggi della tecnologia a stato solido, vi è una contropartita: in generale, i circuiti a stato solido comportano picchi di potenza più bassi di quelli che si possono raggiungere con i radar alimentati da tubi.

Le limitazioni imposte alla potenza del segnale hanno aumentato l'importanza delle tecniche note con il nome di codificazione di impulsi e di compres-

I radar ibridi combinano una schiera in fase con una lente che modifica le caratteristiche del fascio. In una schiera in fase a scansione limitata *(in alto)*, un sistema sempre più usato nei radar di avvicinamento aeroportuali, una piccola schiera in fase, che produce un fascio relativamente largo, emette l'energia all'interno di un riflettore iperbolico. Il fascio è rappresentato per due angoli di puntamento diversi per mostrare come il riflettore lo focalizzi. Il riflettore, pertanto, aumenta il potere risolutivo angolare della schiera, ma ne riduce il campo di visibilità. Questa disposizione permette di usare una schiera più piccola e meno costosa nei casi in cui non è necessario un grande angolo di scansione. In un radar militare sperimentale *(in basso)*, una lente a forma di cupola che copre una schiera orizzontale permette di ottenere l'effetto opposto, aumentando la deflessione del fascio. La lente è costituita o da ceramica speciale o da plastica che rifrange le microonde, o da una configurazione a cupola di sfasatori che incrementano di un valore fisso il ritardo di fase del fascio che ricevono dalla schiera, producendo un riassetto del puntamento a un angolo più marcato. Un radar di questo tipo può esplorare elettronicamente lo spazio dall'orizzonte allo zenit in ogni direzione.

sione di impulsi, tecniche che permettono a un radar di simulare un impulso breve e di alta potenza emettendo segnali meno potenti per un periodo di tempo più lungo. Questa capacità è importante per raccogliere certi tipi di informazione su bersagli distanti, sia che si tratti di riflettori parabolici orientati meccanicamente, sia che si tratti di schiere in fase.

La distanza alla quale un radar di una data sensibilità è in grado di rivelare un oggetto di dimensioni e riflettività particolari dipende dall'energia totale di un impulso. Quanto più breve è l'impulso, tanto più elevata deve essere la potenza di picco di trasmissione dell'impulso, a parità di portata del radar. Il COBRA DANE può rivelare un oggetto metallico delle dimensioni di un pompelmo alla distanza di 1000 miglia nautiche; per ottenere questo risultato con un impulso di cinque nanosecondi, deve irradiare una potenza di picco pari a tremila miliardi di watt, più di quanto basti per distruggere tutti i suoi circuiti.

Eppure, se il radar deve valutare le dimensioni di un oggetto o individuare numerosi oggetti in volo in formazione compatta, è essenziale un impulso molto breve. Il fatto che la portata non dipenda dalla potenza di picco, ma dall'energia totale dell'impulso offre una soluzione: si allunga l'impulso, si riduce in corrispondenza la potenza di picco necessaria con una tecnica chiamata codificazione di impulsi, si comprime poi l'eco in modo da poterne estrarre tutta l'informazione che sarebbe stata fornita da un vero impulso di breve durata. Nel COBRA DANE, per esempio, prima di amplificare e trasmettere un impulso di cinque nanosecondi, lo si allunga di 200 000 volte portandolo a una durata di un millisecondo. La potenza di picco richiesta si riduce di un identico fattore: da tremila miliardi di watt a 15 megawatt, la potenza effettiva del COBRA DANE.

In un tipico procedimento di codificazione di impulsi, un impulso di cinque nanosecondi composto di un intervallo di frequenze passa attraverso una linea di ritardo dispersiva, che ritarda le componenti per valori che aumentano con la frequenza: le frequenze più basse dell'impulso non vengono ritardate, mentre quelle più elevate subiscono il massimo ritardo di un millisecondo. L'impulso, che ora è lungo un millisecondo, viene amplificato e trasmesso; di ritorno viene ricevuta un'eco di uguale durata.

L'eco viene fatta passare attraverso una rete ricevitore-compressore, che aggiunge un secondo insieme di ritardi. In questo caso i ritardi sono in relazione inversa alla frequenza. La frequenza più bassa dell'impulso subisce un ritardo di un millisecondo e la più elevata non subisce alcun ritardo. Nel procedimento di codificazione e compressione degli impulsi, ogni componente del segnale ha subito lo stesso ritardo totale (un millisecondo). Ne risulta un'eco non distorta di cinque nanosecondi.

Se l'impulso trasmesso di un millisecondo, che ha una lunghezza di circa 300 chilometri, colpisce un oggetto sostanzialmente più corto, ne viene riflesso in forma di due echi sovrapposti. Normalmente questi echi non possono venire separati e i loro tempi relativi non possono essere utilizzati per misurare le dimensioni dell'oggetto. Quando però l'eco costituita a codificazione di impulso viene compressa, gli echi sovrapposti si risolvono da sé in due distinti segnali di cinque nanosecondi.

La codificazione e la compressione degli impulsi assolvono funzioni parallele nei radar a stato solido. Anche quando non vi è alcuna necessità di determinare le dimensioni dell'oggetto dal quale il segnale radar viene riflesso, la determinazione accurata della distanza richiede un impulso ragionevolmente breve. Senza la compressione degli impulsi, un impulso di un millisecondo può risolvere la distanza di un oggetto solo a intervalli di 150 chilometri. Inoltre, gli impulsi lunghi sono soggetti a rumori confusi (*clutter*), cioè riflessioni spurie provenienti dal suolo e dalle precipitazioni. E tuttavia le alte potenze richieste per portate adeguate con impulsi di durata minore vanno oltre le possibilità di molti radar a stato solido. La codificazione di impulsi e la compressione di impulsi mettono in grado tali radar di conciliare portata e risoluzione con la bassa potenza.

I primi radar a schiera in fase, installati tra gli anni sessanta e gli anni settanta, sono usati a scopi militari e di spionaggio. In campo civile però vi sono circostanze in cui si presentano le stesse necessità che hanno stimolato lo sviluppo della tecnologia militare: la necessità per esempio di dati frequenti su numerosi oggetti che si muovono velocemente. Un caso del genere si ha vicino all'estremità delle piste degli aeroporti, dove gli aerei in arrivo si mettono in fila per l'avvicinamento finale. In questo caso, i radar di avvicinamento di precisione inviano impulsi agli aerei in arrivo, determinandone contemporaneamente distanza e posizione rispetto all'inizio della pista e guidandoli all'atterraggio. Questi radar utilizzano sempre più la tecnologia delle schiere in fase.

Diminuendo il numero di elementi radianti il costo delle schiere diminuisce. Per la maggior parte degli impieghi, il numero degli elementi radianti è necessariamente alto. Una schiera piccola trasmette un fascio meno focalizzato e quindi più largo. Questo fattore riduce il potere di risoluzione angolare della schiera e la piccola superficie dell'antenna riduce la sensibilità della schiera agli echi. Quando, però, non è necessario esplorare un'ampia zona di cielo, si possono superare entrambi gli inconvenienti di una schiera piccola combinando la schiera con un grande riflettore.

Il campo di visibilità di un radar di avvicinamento di precisione non deve essere necessariamente ampio: un tipico radar di questo genere deve esplorare circa 10 gradi in azimut e da sette a 14 gradi in elevazione. Di conseguenza, un sistema ibrido costituito da una schiera in fase e da un riflettore tradizionale dà buoni risultati. In un sistema di questo tipo una schiera di 443 elementi radianti è combinata con un riflettore parabolico di 4,6 per 4 metri. La schiera è disposta vicino al fuoco del riflettore che riflette il fascio a qualunque angolo di puntamento. Nel funzionamento, il riflettore agisce come una lente, focalizzando il fascio e riducendone la deflessione angolare. I fasci riflessi sono più stretti e sono confinati in un angolo solido più ridotto; di conseguenza, è migliore sia la capacità della schiera di risolvere due bersagli poco distanti angolarmente, sia la sua capacità di determinare la posizione di un singolo bersaglio. Il riflettore parabolico, inoltre, aumenta la sensibilità del radar agli echi.

I futuri radar a schiera in fase usufruiranno dei progressi della tecnologia dei circuiti elettronici. L'introduzione nell'ingegneria radar dell'equivalente dei circuiti integrati digitali permetterà di ridurre considerevolmente il numero e la dimensione dei componenti necessari per generare, ricevere ed elaborare i segnali. Nota con il nome di circuito integrato monolitico a microonde (*monolithic microwave integrated circuitry*, MMIC), questa nuova tecnologia combina sfasatori, commutatori e amplificatori transistorizzati su chip di arseniuro di gallio. Un modulo trasmettitore/ricevitore completo, comprendente tutti i circuiti necessari per un singolo elemento radiante di una schiera in fase, può essere già realizzato con appena 11 chip MMIC. I moduli trasmettitore/ricevitore a stato solido attualmente in uso richiedono invece centinaia di elementi.

In ultima analisi, un modulo di questo tipo può comprendere anche un microelaboratore che, calcolando il giusto ritardo di fase, controlla lo sfasamento, mentre un calcolatore centrale determina l'elevazione e l'azimut del fascio. Nelle attuali schiere in fase questa funzione è devoluta al calcolatore centrale che calcola sequenzialmente le variazioni di fase per le migliaia di elementi radianti. Il calcolo e il controllo della variazione di fase per singoli elementi eliminerebbe le migliaia di cavi attualmente usati per portare i comandi di sfasamento dal calcolatore ai singoli elementi di antenna. Con i progressi della tecnologia circuitale e il procedimento stesso di sfasamento diventerà infine un procedimento digitale, consistente in operazioni matematiche su un segnale digitalizzato. Un tale progresso permetterà di modificare a volontà l'ampiezza del fascio, allargandolo per esplorare vaste zone di cielo o focalizzandolo per l'inseguimento di precisione. Nel frattempo, simili elaborazioni digitali dell'eco potrebbero contribuire a eliminare interferenze dovute a disturbi elettronici intenzionali.

L'aumento di semplicità e di compattezza, insieme alla riduzione di peso e all'aumento di affidabilità, renderanno presto le schiere in fase adatte per compiti spaziali. Schiere in fase orbitanti potrebbero essere utilizzate per realizzare mappe radar della superficie terrestre, controllare il traffico marittimo e il movimento degli iceberg oltre che svolgere compiti militari e di sorveglianza. La prossima generazione di radar per la cartografia della National Aeronautics and Space Administration, trasportati a bordo della navetta spaziale, sarà costituita da schiere in fase; sia la Marina, sia l'Aviazione degli Stati Uniti stanno considerando il lancio di una schiera in fase a bordo di un satellite. Molti dei radar lanciati nello spazio sino a oggi non potevano essere fatti ruotare e quindi nel corso di ciascuna orbita esploravano solo una stretta striscia di superficie terrestre. Il puntamento elettronico, veloce e senza inerzia, di una schiera in fase permetterà invece di tenere larghe strisce della superficie terrestre sotto il controllo del fascio radar.

Il successo riproduttivo del cervo nobile

Uno studio di circa 12 anni condotto su una popolazione di oltre un migliaio di cervi nobili in una riserva naturale scozzese ha consentito di trovare i fattori che determinano il successo riproduttivo di questa specie

di T. H. Clutton-Brock

In una remota valletta dell'isola scozzese di Rhum un maschio di cervo nobile, entrato nel decimo anno d'età, se ne sta ritto in mezzo a un gruppo di femmine. Questo grande harem che lo circonda, sul fitto tappeto erboso, è probabilmente l'ultimo che riuscirà a mantenere. Esso comprende femmine di svariate età. Due matriarche di 15 e 13 anni brucano ai lati opposti del gruppo. Ciascuna presiede a una grande famiglia di figlie mature, con cui condivide un pascolo di due chilometri quadrati, costituito da brughiera, pendii collinosi e fondovalle: il loro «spazio domestico» o «area familiare». Tra loro vi sono le otto figlie, tre con mantello lucido, quarti posteriori arrotondati e senza cerbiatti e cinque accompagnate, invece, dai loro cerbiatti: dalla sagoma spigolosa sono evidenti i segni dell'esaurimento fisico prodotto dall'allattamento. Due femmine di tre anni, che partoriranno per la prima volta nel corso dell'anno, brucano lungo i pendii di una leggera depressione tra le dune. Le otto di due anni, i cerbiatti di un anno e quelli che non l'hanno ancora compiuto sono dispersi nel gruppo. A distanza di meno di 200 metri tre maschi, che hanno già compiuto l'anno d'età e sono stati recentemente separati dai loro gruppi materni dal capo dell'harem, brucano con una certa difficoltà, non essendo ancora del tutto abituati alla vita indipendente.

Sulle collinette sovrastanti appare un cervo di sei anni, il cui manto è totalmente annerito dall'essersi rotolato nelle torbiere. Cammina risolutamente verso l'harem e poi si ferma, emettendo una serie di bramiti intermittenti, come colpi di tosse, più simili al ruggito di un leone che al richiamo d'amore del wapiti, il cervo canadese. Il capo dell'harem replica e per circa 15 minuti i due cervi bramiscono ripetutamente l'uno contro l'altro. Poi, con passi lenti e ponderati, lo sfidante si avvicina ancor più; a questo punto il capo si allontana dal gruppo e si fa incontro al rivale. I due quasi si toccano, ma poi, come per mutuo accordo, si girano, compiendo con andatura nervosa un percorso parallelo di una decina di metri

I contendenti vanno avanti e indietro, finché il cervo residente, individuando un potenziale vantaggio, risale di corsa una piccola altura, abbassa la testa e carica lo sfidante. Il cervo di sei anni, a sua volta, abbassa immediatamente la testa e porta le corna in posizione di scontro: i due contendenti cozzano quindi con un secco rumore, corna contro corna. Entrambi puntano le zampe anteriori in avanti, con i fianchi tesi per spingere con forza. Poi girano per una volta in cerchio. Il più giovane improvvisamente balza in avanti, tentando di spingere di lato il rivale di taglia superiore. Quest'ultimo punta però le zampe posteriori e lentamente obbliga l'avversario a sollevare la testa. Con una rapida torsione gli fa ruotare la testa, lo spinge di lato e lo carica sul fianco esposto. Lo sfidante allora si volta e fugge. Il vincitore lo insegue per alcuni metri; poi si ferma e bramisce contro il rivale in ritirata, prima di tornare verso l'harem. Le femmine non hanno mostrato il benché minimo interesse nei riguardi di questa tenzone e hanno continuato a brucare metodicamente. Il capo manterrà l'harem almeno fino alla prossima sfida: ma alla sua età ha una capacità di lottare sempre più ridotta. Fra un anno la contesa avrebbe probabilmente un esito diverso.

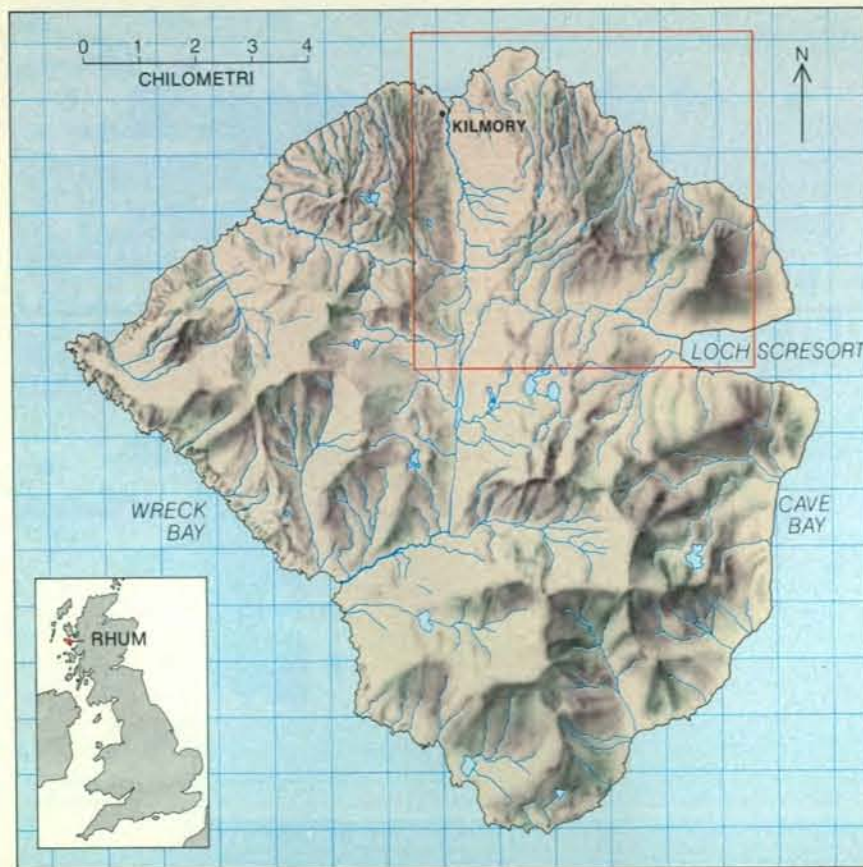
Questa scena evidenzia una diversità fondamentale tra i due sessi, osservabile sostanzialmente in tutti i mammiferi, tra i quali i maschi vincenti hanno il monopolio su parecchie femmine. Poiché maschi e femmine nascono in numero approssimativamente uguale, il fatto che alcuni maschi siano in grado di mantenere harem di 10-20 femmine ha come necessaria conseguenza che ogni anno un'elevata percentuale di altri maschi non ha la possibilità di riprodursi. Pertanto la competizione per gli harem è intensa e i maschi combattono regolarmente per il possesso delle femmine.

La situazione è diversa nel caso delle femmine. Il successo riproduttivo di una femmina non è in rapporto diretto con il numero di partner, ma dipende dal numero di piccoli che può partorire e allevare nel corso dell'esistenza. Poiché il dispendio di energie per la gestazione e l'allattamento è rilevante, è probabile che il successo di una femmina dipenda soprattutto dalle risorse di cibo che ha a disposizione.

Poiché nelle specie in cui i maschi possiedono degli harem i due sessi si comportano in modo completamente diverso nella competizione, anche i fattori che influiscono sul successo riproduttivo debbono essere diversi. Per i maschi sono importanti l'abilità nel combattere, le dimensioni corporee imponenti e le armi molto sviluppate. Le femmine, invece, avrebbero ben poca convenienza a combattere per risorse alimentari molto disperse: perciò il loro successo dovrebbe dipendere soprattutto dall'efficienza con la quale ricavano cibo dall'ambiente e lo utilizzano per il perpetuarsi della specie. Con ciò non si vuol dire che la loro capacità nel prendere il posto di altre femmine nei terreni di pascolo non abbia importanza; tuttavia, l'abilità nel combattimento non ha probabilmente, a differenza dei maschi, un effetto significativo sul successo riproduttivo.

Le differenze nei fattori che determi-

Un maschio di cervo nobile è in compagnia del suo harem sull'isola di Rhum, nelle Ebridi interne. Il successo riproduttivo dei maschi di questa specie aumenta con le dimensioni dell'harem. Queste, a loro volta, dipendono dall'abilità del maschio in combattimento: un maschio sconfitto da uno sfidante è privato dell'harem e, di solito, non riesce più a riconquistarlo. In mezzo al gruppo è riconoscibile un cerbiatto. Le cinque femmine stanno cercando alghe marine lungo la costa. Il numero di femmine in un harem può variare da uno a 20.

Il capo dell'harem risponde alla sfida lanciata da un maschio più giovane con un potente bramito. Se lo sfidante continua ad avvicinarsi, i due maschi camminano in parallelo, così da potersi valutare a vicenda da vicino. Circa il 40 per cento di queste «passeggiate» porta a un combattimento.

L'isola di Rhum è la località in cui, nel corso di 12 anni, è stato compiuto da parte dell'autore e collaboratori uno studio sul cervo nobile. L'isola, situata al largo della costa nordoccidentale della Scozia, è una riserva naturale del Regno Unito. L'area di studio, nella quale si trovano in media circa 300 cervi, è situata nella parte nordorientale dell'isola *(riquadro in rosso)*.

nano il successo riproduttivo nei maschi e nelle femmine sono probabilmente più nette nelle specie poligine (cioè in quelle specie in cui i maschi monopolizzano il legame riproduttivo con parecchie femmine) e nelle specie poliandre (in cui sono le femmine che monopolizzano il legame riproduttivo con parecchi maschi) rispetto alle specie monogame. In queste ultime la competizione tra maschi e femmine assume probabilmente aspetti più simili.

I fattori che influiscono sul successo riproduttivo, e che sono diversi nei maschi e nelle femmine, sono la causa principale della maggior parte delle differenze tra i due sessi. Se si eccettuano gli animali che si riproducono comunitariamente, la selezione favorirà i caratteri che aumentano il successo riproduttivo del singolo individuo. Quando un carattere influisce in misura maggiore sul successo riproduttivo di uno dei due sessi è probabile che sia più sviluppato in quel sesso. Per esempio, nelle specie in cui le dimensioni corporee influiscono sul successo riproduttivo dei maschi più che su quello delle femmine e il dispendio di energie negli animali di grossa taglia è uguale nei due sessi, i maschi sono di dimensioni maggiori di quelle delle femmine. Il fenomeno è evidente nel cervo nobile e nel wapiti i cui maschi hanno un peso quasi doppio delle femmine. Come logica conseguenza di questo ragionamento, le differenze sessuali generalmente aumentano con il grado di poliginia o di poliandria.

Se alla radice della maggior parte delle differenze sessuali riguardanti la morfologia e il comportamento vi sono differenze nei fattori che determinano il successo riproduttivo, la conoscenza di questi fattori nei maschi e nelle femmine dovrebbe servire a formulare previsioni e a spiegare le differenze tra i sessi. Tuttavia non è facile scoprire le cause di tale variazione. Ciò che conta, naturalmente, è il successo dell'individuo nel corso di tutta l'esistenza. Eppure anche il numero di cuccioli allevati può non costituire un buon metodo di valutazione del successo, poiché i piccoli nati in cucciolate numerose possono avere una minore probabilità di riprodursi bene. Una valutazione migliore del successo riproduttivo di un individuo si otterrebbe dal computo dei nipoti: tuttavia, i problemi logistici associati all'accertamento preciso di questo dato tra gli animali selvatici sono enormi. Non sono molte le specie in cui è possibile valutare il successo riproduttivo dei singoli individui addirittura nell'arco di una stagione: figuriamoci, quindi, calcolare il numero di nipoti di ciascun individuo!

Un'eccezione a questa regola è costituita dal cervo nobile o cervo rosso *(Cervus elaphus)* che vive in Scozia, considerato di solito appartenente al medesimo genere del wapiti, il cervo canadese. L'osservazione di ambedue i sessi di questa specie è facile nell'ambiente degli Highlands scozzesi, praticamente privo di alberi. Con un po' di pratica, un osservatore può imparare a riconoscere i singoli individui dalla struttura del muso e dal disegno del mantello. I cervi hanno una stagione riproduttiva breve e prevedibile, in ottobre, quando i maschi adulti si pongono a difesa di harem ben definiti. I cerbiatti, che nascono nel giugno successivo, possono essere in genere catturati, marcati e pesati. Inoltre, entrambi i sessi hanno una durata di vita abbastanza breve (circa 10 anni i maschi, 12 le femmine) e quindi i singoli individui possono essere seguiti per tutto il ciclo vitale.

Per queste ragioni, con i miei collaboratori Fiona Guinness e Stephen Albon dell'Università di Cambridge, studio da 12 anni i singoli individui di cervo nobile nell'isola di Rhum, che fa parte delle Ebridi interne. Nella nostra zona di studio si trovano in media circa 300 cervi; tuttavia, nel corso dei 12 anni in cui si è prolungata l'osservazione, ben 1000 cervi hanno fatto parte della popolazione. Negli anni recenti, i cerbiatti catturati e marcati durante le prime fasi del nostro studio hanno cominciato a completare il ciclo vitale: è adesso quindi che comincia a emergere un quadro generale di ciò che determina il successo riproduttivo nei maschi e nelle femmine.

Il successo riproduttivo varia più tra i maschi adulti che tra le femmine. Il numero di cerbiatti che un maschio genera ogni anno varia da 0 a circa 10. (Questa cifra scende a 6 quando si considerano solo i cerbiatti che sopravvivono per almeno un anno).

Le differenze di successo sarebbero ben più evidenti se queste cifre contrastanti rimanessero costanti per tutta la vita di un maschio, ma ciò non accade. Il successo riproduttivo è fortemente influenzato dall'età e, di solito, ha un massimo tra i sette e i nove anni. In un dato momento è possibile stimare il successo riproduttivo di un maschio in base alla dimensione del suo harem; tuttavia, se consideriamo tutta la durata dell'esistenza di quel maschio, la variazione nel successo riproduttivo è minore di quanto farebbero pensare le notevoli differenze nelle dimensioni dell'harem osservabili in una singola stagione. Persino i maschi con maggiore successo riproduttivo, nel corso della loro esistenza, di rado generano più di una ventina di cerbiatti che superano l'anno d'età.

Invece, la variazione nel successo riproduttivo delle femmine è più rilevante di quanto lasci supporre il fatto che una femmina partorisce al massimo un cerbiatto all'anno. La ragione è che le stesse femmine tendono a perdere i piccoli durante tutto il periodo fecondo (che varia dai sei ai 12 anni). Il numero dei cerbiatti sopravvissuti, che le femmine riescono ad allevare nel corso dell'esistenza, varia così da 0 a 10 ed è in stretto rapporto con il numero dei nipoti.

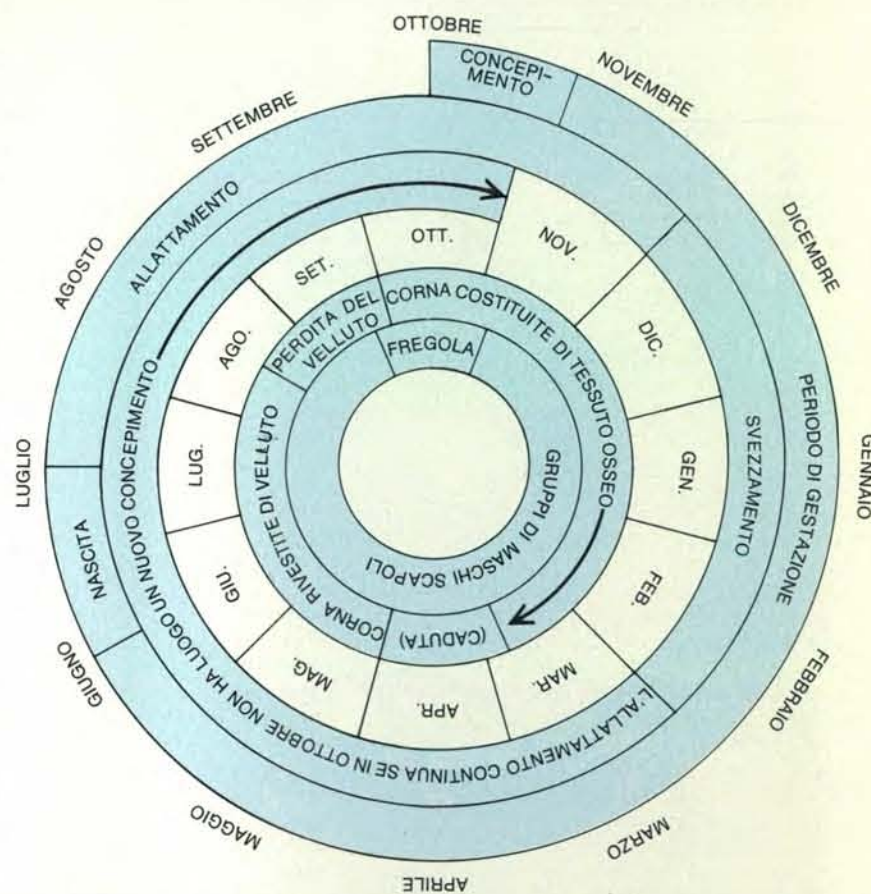
Il successo riproduttivo di un individuo consta, dunque, di tre componenti: la durata dell'esistenza, il numero di piccoli procreati ogni anno e la percentuale di quelli che riescono a sopravvivere. L'importanza relativa di queste componenti è notevolmente diversa nei due sessi. Il successo riproduttivo di una femmina dipende in parte dalla lunga durata della vita, ma prevalentemente dalla percentuale dei cerbiatti che riescono a sopravvivere. Alcune femmine partoriscono un cerbiatto all'anno per quasi tutta la durata del periodo fecondo, ma non riescono a farne crescere a lungo neppure uno.

Tra i maschi, invece, le differenze di successo riproduttivo dipendono quasi esclusivamente dal numero di cerbiatti che procreano ogni anno tra l'età di sei e di 11 anni. Le probabilità di sopravvivenza dei cerbiatti generati dai diversi cervi variano di poco e la durata della vita di un maschio oltre gli 11 anni ha conseguenze trascurabili sul successo riproduttivo considerato globalmente per tutta la sua esistenza.

Anche i fattori che influiscono sul successo riproduttivo sono diversi tra maschi e femmine. La capacità di una femmina di allevare la prole è influenzata soprattutto dalla qualità dell'area familiare o spazio domestico (da cui dipende la possibilità di accedere alle risorse alimentari), dal numero di animali con cui convive (da cui dipende il grado di competizione per il cibo) e dal livello di dominanza sociale (che condiziona la sua priorità nell'accedere ai luoghi preferiti per il pascolo). Una femmina eredita dalla madre l'area familiare e la condivide con le sorelle e con gli altri parenti della linea materna: perciò il suo successo riproduttivo dipende in genere dall'identità e dalle dimensioni della famiglia a cui appartiene. Il suo rango di dominanza è in rapporto con il peso alla nascita e in età adulta: le femmine con notevole peso alla nascita crescono rapidamente e si sviluppano in adulte di grande taglia e dominanti.

Al contrario il successo riproduttivo dei maschi dipende quasi interamente dalla capacità di combattere. Questa, a sua volta, è in rapporto con le dimensioni corporee e con la crescita durante il primo anno di vita. Nel cervo nobile come in molte altre specie in cui i maschi sono notevolmente più grandi delle femmine, la crescita dei maschi è influenzata da una carenza di cibo in modo più grave di quella delle femmine. In parte per questa ragione e in parte perché varia più ampiamente nei maschi che nelle femmine, il successo riproduttivo è influenzato nei maschi dal grado di dominanza e dalle dimensioni della madre in modo ancor più netto che nelle femmine. Di conseguenza, i figli maschi delle femmine dominanti hanno più successo delle figlie, mentre le figlie delle femmine subordinate superano i figli maschi delle stesse.

I fattori opposti che influiscono sul successo riproduttivo dei maschi e delle femmine di cervo nobile, permettono di intuire le origini evolutive di molte differenze sessuali nella morfologia, nella fisiologia e nel comportamento. Le ragioni di talune differenze sessuali sono ovvie. Il fatto che le dimensioni e la capacità di lottare siano così importanti nei maschi del cervo nobile spiega perché questi animali hanno un peso quasi doppio di quello delle femmine e hanno sviluppato armi così elaborate come le tipiche corna ramificate o palchi. Come è implicito in questa interpretazione, le differenze sessuali nelle dimensioni corporee sono massime in quelle specie di cervidi, come per esempio renne e alci, in cui i maschi con un alto successo riproduttivo mantengono un harem numeroso. Tali differenze sono, invece, minime in specie come i caprioli e i cervuli Muntjak, nelle quali la possibilità, per i maschi, di accoppiamenti poligini (o poligami) è scarsa a causa dell'ampia distribuzione delle femmine.

I fattori opposti che influiscono sul successo riproduttivo dei maschi e delle femmine forniscono anche chiari indizi sull'origine di alcune differenze meno ovvie tra i sessi. In molte specie di mammiferi (compreso il cervo nobile) i maschi giovani crescono più rapidamente delle femmine, ma accumulano nel corpo minori riserve di grasso, rinunciando ai vantaggi forniti da un accumulo adiposo a favore dei vantaggi forniti,

Nella rappresentazione del ciclo riproduttivo del cervo nobile sull'isola scozzese di Rhum, i due cerchi esterni si riferiscono al ciclo delle femmine, i due interni a quello dei maschi. L'epoca della riproduzione è in autunno e oltre l'80 per cento delle femmine concepisce in ottobre. Cominciando a seguire l'andamento del ciclo femminile in ottobre *(in alto)*, si osserva che nel giugno successivo nasce il cerbiatto: la maggior parte delle nascite ha luogo appunto in questo mese. Generalmente i piccoli vengono svezzati dopo sei o sette mesi se la madre concepisce di nuovo; se il concepimento non ha luogo nell'autunno dell'anno successivo, allora l'allattamento si protrae per circa 18 mesi a partire dalla nascita. Il cerchio esterno, dei due relativi ai maschi, indica lo stato delle corna mese per mese; l'altro evidenzia con chiarezza la stagione riproduttiva del cervo nobile, che si svolge sostanzialmente in ottobre.

invece, da una rapida e precoce crescita e da dimensioni corporee maggiori nell'età adulta.

Anche queste differenze variano sistematicamente da specie a specie. Quelle relative all'adiposità giovanile sono più pronunciate nelle specie poligine (in cui anche altre differenze fisiche si esprimono nella maniera più evidente). Nelle specie con minor dimorfismo l'adiposità degli animali giovani, maschi e femmine, è all'incirca uguale. Nell'idropote inerme o capriolo d'acqua cinese, l'unica specie di cervide le cui femmine sono di dimensioni maggiori rispetto ai maschi, i maschi giovani sono più grassi delle femmine giovani.

Caratteristiche legate al sesso maschile, come maggiore mole, maggiore tasso di crescita e ridotte riserve di grasso, devono però avere un costo, altrimenti le differenze tra i sessi continuerebbero ad aumentare. Nel cervo nobile e in molti altri mammiferi caratterizzati da dimorfismo sessuale, i maschi hanno vita più breve delle femmine e maggiori probabilità di morire di fame durante i periodi di carestia o di inclemenza del tempo. Uno degli esempi più spettacolari a questo proposito è stato quello della popolazione di renne sull'isola di San Matteo nel Mare di Bering: una popolazione di oltre 6000 adulti scese a soli 42 nel corso di un singolo inverno, e solo uno degli individui sopravvissuti era un maschio.

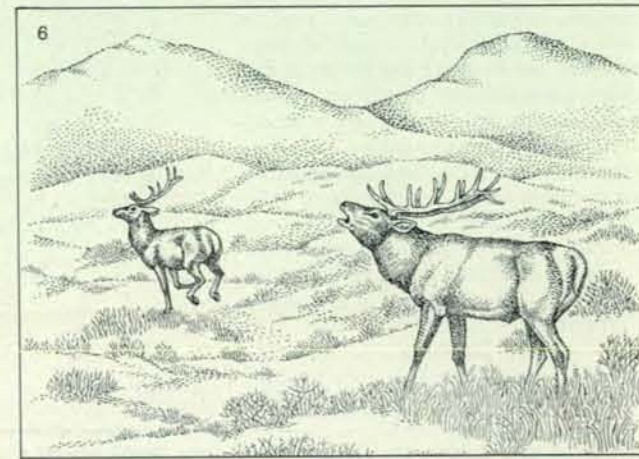
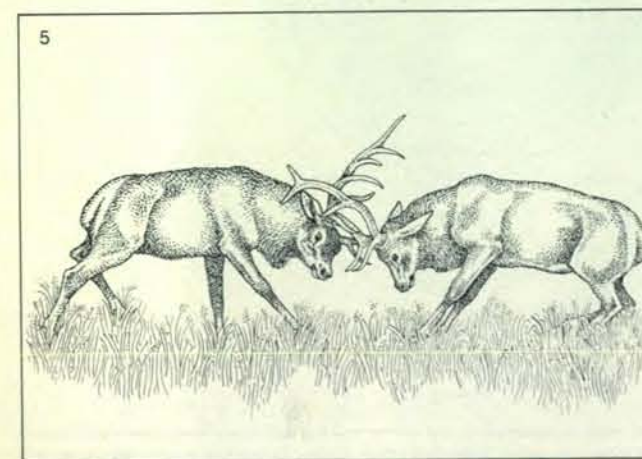
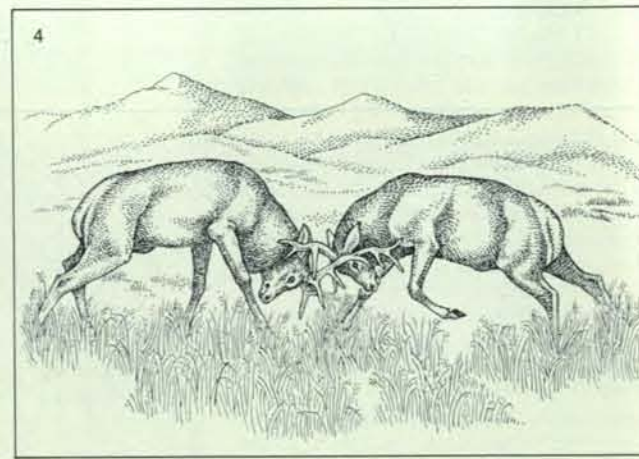
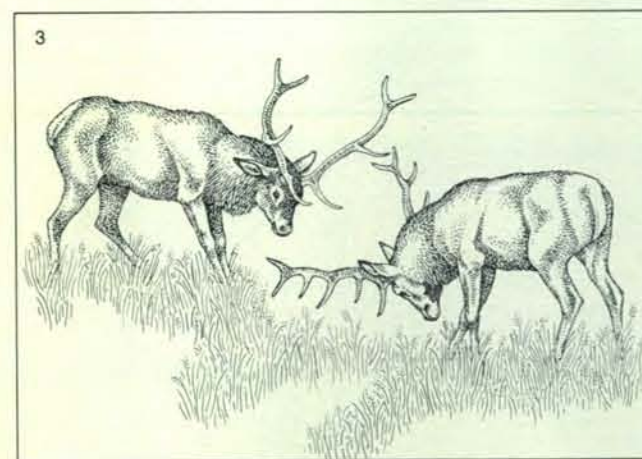
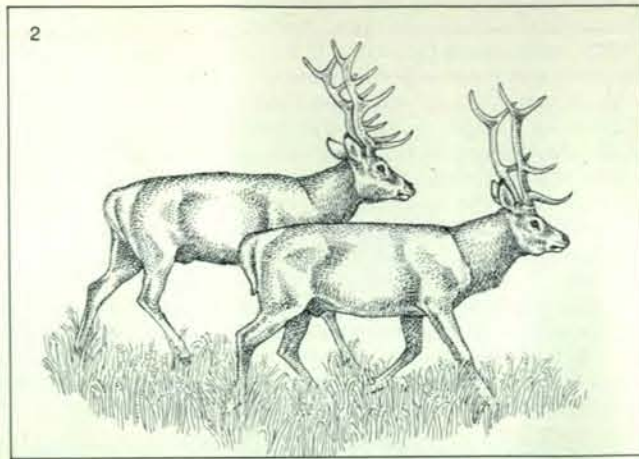
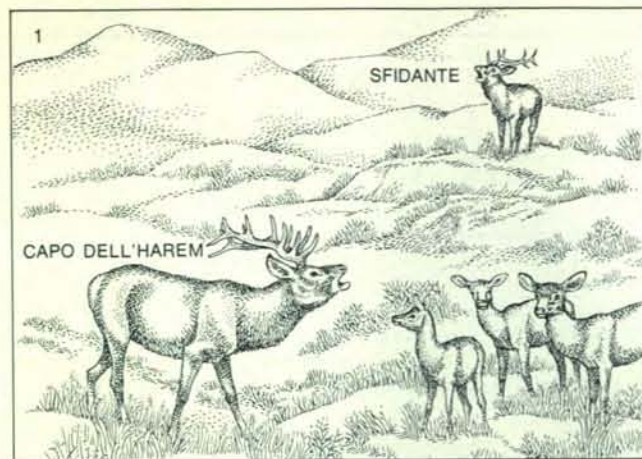
Anche tra individui giovani si trovano differenze sessuali nel grado di sopravvivenza. I maschi di cervo nobile, all'età di meno di due anni, hanno molte più probabilità di morire durante l'inverno rispetto alle femmine. Nei primi anni della nostra indagine, quando la densità della popolazione era piuttosto bassa, non apparvero differenze sessuali nella mortalità giovanile, ma quando la densità aumentò la situazione cambiò. Oggi circa il 60 per cento dei maschi muore durante il primo o il secondo inverno. Per le femmine l'incidenza è del 30 per cento.

Le differenze sessuali riguardo agli effetti che una crescita precoce ha sul successo riproduttivo fanno sorgere il dubbio che, nei mammiferi, le madri possano trattare figli e figlie in modo diverso. Il fatto poi che la crescita precoce e l'investimento materno sotto forma di latte abbiano probabilmente un'influenza sul successo riproduttivo dei maschi suggerisce la possibilità che le femmine aumentino il numero dei nipoti dedicando la maggior parte del loro impegno nella riproduzione ai propri figli maschi che non alle figlie. Esistono alcune prove che esse si comportino proprio così. I maschi del cervo nobile nascono con un peso superiore di circa mezzo chilogrammo rispetto alle femmine. Differenze simili si osservano in specie che mettono al mondo figliate.

Dato che la crescita durante le prime settimane di vita ha la probabilità di influenzare il successo riproduttivo dei maschi più di quello delle femmine, è logico aspettarsi che le madri forniscano più latte ai figli maschi. I cerbiatti maschi di cervo nobile sull'isola di Rhum poppano più a lungo e più spesso delle femmine. In parecchie altre specie di mammiferi poligini è stata osservata una maggiore intensità di allattamento proprio nei riguardi dei piccoli di sesso maschile.

Il confronto che abbiamo effettuato su femmine a seconda che avessero allevato figli maschi o femmine dimostra che l'allevamento di maschi comportava per le madri i maggiori costi in termini di un ulteriore successo riproduttivo. In qualsiasi stagione circa il 25 per cento delle femmine non riesce a mettere al mondo cerbiatti. Questa incapacità è in larga misura una conseguenza delle condizioni fisiche scadenti durante l'autunno. Nella popolazione da noi osservata il 33 per cento delle madri che allevano figli maschi non riesce a riprodursi nel corso dell'anno successivo, in confronto al 18 per cento delle madri che allevano femmine. Anche la data del concepimento ha un rapporto con le condizioni della femmina in autunno. Nell'anno successivo a quello in cui hanno allevato un maschio le femmine concepiscono circa due settimane più tardi rispetto all'anno successivo a quello in cui hanno allevato una femmina.

La tesi secondo cui le madri dovrebbero investire maggiormente in figli maschi per aumentare il numero dei nipoti non si applica necessariamente a tutti i membri di una popolazione. Come per primo ha dimostrato Robert L. Trivers dell'Università della California a Santa Cruz, le madri che a causa della debole costituzione fisica hanno scarse probabilità di allevare figli maschi sufficientemente dotati, potrebbero abbandonare i piccoli di sesso maschile appena dopo la nascita, risparmiando così le proprie risorse per successivi tentativi. Per esempio, queste femmine subordinate possono persino arrivare a negare il latte ai figli maschi che presumibilmente non potranno riprodursi con successo.

È proprio vero tutto questo? Se la risposta è affermativa, i figli maschi delle madri subordinate avrebbero mag-

La sfida al capo di un harem ha inizio quando un giovane maschio gli si avvicina *(1)*. I due contendenti di solito si lanciano vicendevolmente bramiti minacciosi per parecchi minuti. Se lo sfidante si avvicina ulteriormente, il capo dell'harem esce dal branco e lo affronta. Nella maggioranza dei casi la mossa successiva è una camminata parallela che consente ai due animali di studiarsi a vicenda *(2)*. Entrambi i cervi avanzano e indietreggiano assieme, tenendosi alla distanza di alcuni metri, per parecchi minuti. Se l'intruso non si ritira, uno dei due contendenti inizia il combattimento girando e abbassando le corna *(3)*. I due animali congiungono le corna e spingono con forza *(4)*. Nello scontro qui raffigurato il capo dell'harem conclude vittoriosamente il combattimento, mentre il rivale è costretto a ritirarsi *(5)*. Al termine dell'incontro, lo sfidante si dà alla fuga, mentre il vincitore lo insegue per alcuni metri, continuando a bramire minacciosamente contro di lui *(6)*.

ANNI	FEMMINE NATE NEL 1972		
	FEMMINA N. 1 (THII)	FEMMINA N. 2 (MOMC)	FEMMINA N. 3 (COLC)
1972	THII (NEONATA)	MOMC (NEONATA)	COLC (NEONATA)
1973	THII (UN ANNO DI ETÀ)	MOMC (UN ANNO DI ETÀ)	COLC (UN ANNO DI ETÀ)
1974	THII (DUE ANNI DI ETÀ)	MOMC (DUE ANNI DI ETÀ)	COLC (DUE ANNI DI ETÀ)
1975	NESSUNA NASCITA	NESSUNA NASCITA	IL CERBIATTO SOPRAVVIVE
1976	IL CERBIATTO SOPRAVVIVE	IL CERBIATTO MUORE IN ESTATE	IL CERBIATTO MUORE IN ESTATE
1977	IL CERBIATTO SOPRAVVIVE	IL CERBIATTO MUORE IN ESTATE	IL CERBIATTO SOPRAVVIVE
1978	THII MUORE	IL CERBIATTO MUORE IN ESTATE	IL CERBIATTO SOPRAVVIVE
1979		IL CERBIATTO MUORE NEL SECONDO INVERNO	IL CERBIATTO SOPRAVVIVE
1980		IL CERBIATTO MUORE NEL PRIMO INVERNO	IL CERBIATTO SOPRAVVIVE
1981		NESSUNA NASCITA	IL CERBIATTO MUORE IN ESTATE
1982		IL CERBIATTO MUORE NEL SECONDO INVERNO	IL CERBIATTO SOPRAVVIVE
1983		IL CERBIATTO MUORE NEL PRIMO INVERNO	IL CERBIATTO SOPRAVVIVE
1984		NESSUNA NASCITA	NESSUNA NASCITA
TOTALE DEI CERBIATTI SOPRAVVISSUTI	2	0	7

Per illustrare il successo riproduttivo del cervo nobile sono state prese in esame tre delle numerose femmine nate nel 1972 sull'isola di Rhum. Confrontando Momc con Colc si osserva che il successo riproduttivo, misurato come numero di cerbiatti sopravvissuti fino all'età adulta, varia notevolmente tra i vari soggetti. Il rango sociale è un fattore importante nel determinare il successo riproduttivo: infatti Colc è relativamente dominante, mente Momc è subordinata.

giori probabilità di morire delle femmine, mentre non si dovrebbero rilevare differenze sessuali nella sopravvivenza tra figli maschi e femmine di femmine dominanti. I dati riferibili al cervo nobile mostrano esattamente questo quadro: il 51 per cento dei maschi nati da madri subordinate muore entro i primi due anni di vita, contro il 30 per cento delle femmine. Invece tra figli maschi e femmine di madri dominanti non si osservano differenze di questo tipo: il 38 per cento dei maschi muore entro i primi due anni, in confronto al 36 per cento delle femmine.

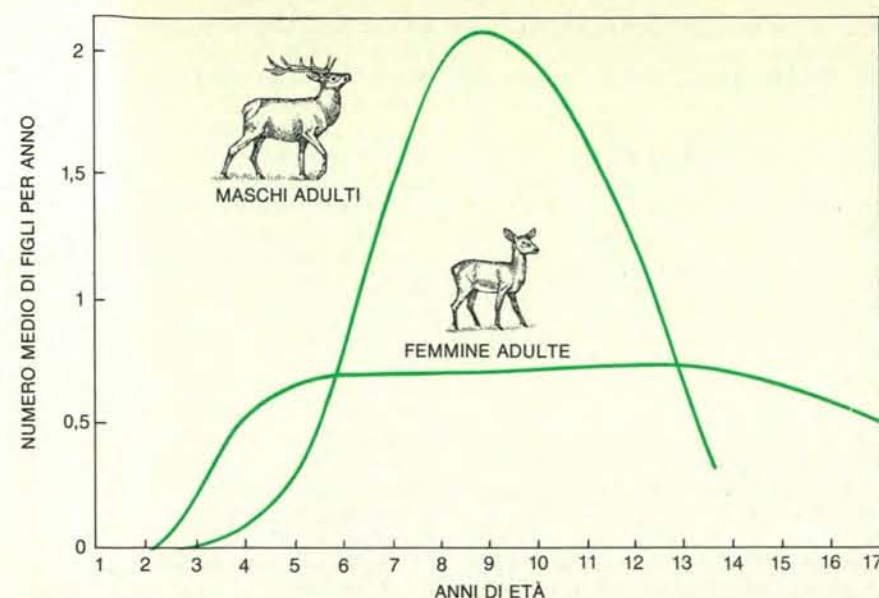
Ma anche l'abbandono della prole appena nata comporta un onere gravoso: ogni cerbiatto rappresenta, infatti, una frazione rilevante del potenziale riproduttivo di una femmina per tutto il corso della sua esistenza. In teoria, le madri potrebbero aumentare il proprio successo riproduttivo modificando il rapporto numerico dei sessi nella prole prima della nascita, idealmente il più vicino possibile al concepimento. La relazione tra rango della femmina e successo riproduttivo della prole, maschi e femmine, lascia pochi dubbi sul fatto che le madri dominanti vedrebbero aumentare il numero dei nipoti mettendo al mondo più maschi che femmine e che il contrario avrebbe luogo per le madri subordinate.

Tutto questo accade veramente? Abbiamo provato a confrontare femmine situate in tre categorie uguali di rango sociale. Esaminando il rapporto numerico tra i due sessi nei cerbiatti si nota il 47 per cento circa di maschi nel caso di femmine subordinate, il 54 per cento nel caso di femmine di rango sociale intermedio, il 61 per cento nel caso di femmine dominanti. Le differenze sono costanti in tutte le otto coorti di femmine sulla cui riproduzione possediamo dati, riguardanti almeno il 75 per cento del ciclo vitale. Inoltre, considerando le femmine una per una, il rapporto numerico tra i due sessi dei cerbiatti aumenta con il crescere del rango sociale e passa dal 30 per cento circa al 70 per cento circa nel caso di femmine dominanti. Il rango in cui le madri superano, con la figliata, il rapporto numerico medio tra i sessi della popolazione intera coincide perfettamente con il punto in cui i figli maschi presentano un successo riproduttivo maggiore di quello delle femmine. Uno studio condotto a Portorico sui bunder ha dato risultati del tutto analoghi.

Questi risultati sono sorprendenti, perché nelle ricerche condotte su uccelli e mammiferi addomesticati si è constatato che è impossibile manipolare il rapporto numerico tra i sessi alla nascita mediante tecniche di incrocio selettivo. Una possibile spiegazione è che una variazione costante nel rapporto numerico tra i sessi può aver luogo in risposta a differenze individuali nelle dimensioni o nelle condizioni fisiche della madre in popolazioni in cui le risorse alimentari sono limitate. Forse gli animali addomesticati non mostrano tendenze corrispondenti ai dati precedentemente rilevati perché vengono mantenuti a un livello di nutrizione artificialmente elevato: le differenze nelle condizioni fisiche sono, pertanto, piccole. La capacità di variare il rapporto numerico tra i due sessi in risposta a variazioni ambientali potrebbe anche essere andata perduta tra gli animali addomesticati per effetto della selezione artificiale praticata su molte generazioni.

Nella nostra ricerca non siamo ancora riusciti a dimostrare se la variazione nel rapporto numerico tra i sessi, derivante da differenze nella mortalità postnatale o da una variazione nel rapporto numerico tra maschi e femmine alla nascita, sia controllata dalla madre. Può darsi che i giovani maschi siano in genere più suscettibili alle carenze alimentari delle giovani femmine. Se così fosse, si spiegherebbe come mai i figli maschi di madri subordinate hanno maggiori probabilità di morire rispetto alle femmine. I confronti compiuti tra l'allattamento dei cerbiatti maschi e quello delle femmine di madri subordinate hanno per ora fornito risultati ambigui.

Può darsi anche che le femmine subordinate partoriscano un minor numero di figli maschi rispetto alle dominanti perché la progenie di sesso maschile ha maggiori probabilità di morire prima della nascita. Negli studi compiuti su diversi mammiferi di vario tipo si è potuto dimostrare che la percentuale dei maschi, tra i feti abortiti, è maggiore di quella delle femmine.

La ridotta sopravvivenza dei feti di sesso maschile spiegherebbe, tuttavia, i nostri risultati solo se tutte le femmine concepissero con un rapporto numerico a favore del sesso maschile. In verità, le femmine dominanti partoriscono figli maschi in un numero nettamente superiore al 50 per cento del totale. Inoltre, se la tendenza osservata appare perché le madri subordinate perdono i feti di sesso maschile durante la gestazione, queste madri dovrebbero partorire meno frequentemente delle dominanti. Invece le differenze di fertilità tra dominanti e subordinate sono troppo piccole per giustificare la differenza osservata nel rapporto numerico tra i sessi dei cerbiatti neonati.

ANNI	MASCHI NATI NEL 1972			
	MASCHIO N. 1 (BROC)		MASCHIO N. 2 (YHOC)	
	DIMENSIONI MEDIE DELL'HAREM	CERBIATTI PROCREATI	DIMENSIONI MEDIE DELL'HAREM	CERBIATTI PROCREATI
1972	BROC NEONATO		YHOC NEONATO	
1973	BROC (UN ANNO DI ETÀ)		YHOC (UN ANNO DI ETÀ)	
1974	BROC (DUE ANNI DI ETÀ)		YHOC (DUE ANNI DI ETÀ)	
1975	BROC (TRE ANNI DI ETÀ)		YHOC (TRE ANNI DI ETÀ)	
1976	NESSUN HAREM	NESSUNA NASCITA	UNA FEMMINA	NESSUNA NASCITA
1977	NESSUN HAREM	NESSUNA NASCITA	QUATTRO FEMMINE	QUATTRO NATI
1978	UNA FEMMINA	NESSUNA NASCITA	TRE FEMMINE	DUE NATI
1979	NESSUN HAREM	NESSUNA NASCITA	QUATTRO FEMMINE	DUE NATI
1980	TRE FEMMINE	UN NATO	CINQUE FEMMINE	TRE NATI
1981	DUE FEMMINE	NESSUNA NASCITA	CINQUE FEMMINE	DUE NATI
1982	NESSUN HAREM	NESSUNA NASCITA	YHOC MUORE	
1983	BROC MUORE			

Il successo riproduttivo di due maschi di cervo nobile è studiato in termini di dimensioni dell'harem e numero stimato di cerbiatti procreati ogni anno. L'entità numerica dell'harem dipende dalla capacità del maschio di uscire vittorioso dai combattimenti contro tutti gli sfidanti: questa capacità gli garantisce il dominio su un certo numero di femmine e raggiunge il massimo a un'età compresa tra i sette e i dieci anni.

Il successo riproduttivo dei maschi di cervo nobile è influenzato dall'età. Il maschio migliora la capacità di combattere a mano a mano che invecchia; ma poi, passato il periodo migliore, verso i 10 anni, questa capacità declina rapidamente. Tra le femmine, invece, l'età non è tanto importante, poiché il successo riproduttivo non dipende dalla capacità di combattere.

Le nostre osservazioni sui cicli vitali dei cervi nobili dell'isola di Rhum sollevano molte questioni riguardanti le cause delle tendenze riproduttive nei mammiferi: a questi interrogativi si potrà rispondere soltanto mediante esperimenti controllati in laboratorio. Perché i maschi sono più suscettibili alle carenze alimentari delle femmine? Le madri reagiscono realmente in modo diverso nei confronti dei piccoli di sesso maschile e femminile, oppure le differenze nell'intensità dell'allattamento sono il risultato di differenze sessuali nella richiesta di latte? Quali meccanismi sono responsabili del rapporto numerico a favore dei maschi nella nascita dei piccoli di molti mammiferi? Le femmine subordinate di piccola taglia concepiscono più figli maschi rispetto alle dominanti?

È difficile riuscire a fornire risposte definitive a questi interrogativi considerando esclusivamente l'isola di Rhum. Cionondimeno, la nostra ricerca serve a chiarire il ruolo che gli studi sulle strategie riproduttive degli animali nel loro ambiente naturale possono avere e suggerisce l'indirizzo di ricerca in settori affini. Per citare un esempio, sembra probabile che i tentativi di manipolare sperimentalmente il rapporto numerico tra i sessi alla nascita avrebbero potuto avere maggiore successo se fossero stati compiuti su animali dotati di un pronunciato dimorfismo sessuale e mantenuti in condizioni di alimentazione limitata.

Allo scopo di verificare molte delle spiegazioni funzionali suggerite a riguardo delle differenze sessuali, gli studi futuri dovranno esaminare i fattori determinanti il successo riproduttivo nelle specie che non si conformano alle tendenze generali. Per esempio, i maschi delle zebre e delle iene macchiate, come i cervi nobili, sono poligini: eppure le differenze di dimensioni tra maschi e femmine nelle zebre sono minime, mentre le femmine delle iene macchiate sono addirittura più grandi dei maschi. L'effetto delle dimensioni corporee sul successo riproduttivo è simile nei maschi e nelle femmine delle zebre? Le dimensioni influiscono sul successo riproduttivo delle femmine della iena macchiata più che su quello dei maschi?

Sono anche necessari studi sulle specie che presentano sistemi riproduttivi opposti. I fattori determinanti il successo riproduttivo in entrambi i sessi sono in generale più simili nelle specie monogame di quanto non lo siano nelle specie poligine? Come differiscono nelle specie poliandre? La sfida che oggi si pone ai biologi sul campo che si occupano di questi argomenti, sarà quella di trovare specie in cui è possibile misurare il successo riproduttivo di entrambi i sessi per tutta la durata del ciclo vitale naturale e di ricercare le cause di diversità tra i singoli individui.

(RI)CREAZIONI AL CALCOLATORE

di A. K. Dewdney

Un sistema esperto batte i semplici mortali nella conquista dei terribili Sotterranei del Destino

Ogni anno, migliaia di persone incontrano una morte di fantasia nei Sotterranei del Destino giocando ad Avventuriero (Rogue, in inglese), un rappresentante di una nuova generazione di giochi d'avventura con il calcolatore. Il giocatore osserva una mappa dei sotterranei sullo schermo e dirige le azioni di un personaggio chiamato l'Avventuriero. Obiettivo del gioco è scendere attraverso i 26 livelli dei sotterranei, impadronirsi dell'Amuleto di Yendor e tornare sani e salvi in superficie raccogliendo quanto più oro possibile lungo il percorso e uccidendo o sfuggendo ai mostri che si incontrano. Sono ben pochi i giocatori umani che riescono, non dico a tornare con l'amuleto, ma anche solo a sopravvivere ai pericoli di questa odissea sotterranea.

La sensazione di entrare davvero in un mondo fantastico può essere molto vivida: è facile lasciarsi rapidamente alle spalle la coscienza di essere davanti a una tastiera e a uno schermo. Come l'Avventuriero, mi avvicinai all'ingresso dei sotterranei con una certa trepidazione: era notte e le antiche rovine che indicavano il punto della mia imminente discesa avevano un aspetto lugubre e minaccioso. Preparai la mia mazza incantata, un arco e una faretra di frecce sottratti dal tesoro di un drago sulle Montagne Oscure. Indossai la mia armatura costruita dai folletti, afferrai le mie armi e i viveri e mi inoltrai nell'oscurità stigea di una scala.

Proprio quando la discesa cominciava ad apparirmi senza fine, mi trovai di fronte a una porta di quercia e la aprii con cautela. Davanti a me c'era la prima stanza del livello superiore dei Sotterranei del Destino. Delle candele spandevano una fievole luce e io mi spinsi fino al centro della camera, per sorvegliarla meglio. Improvvisamente, il pavimento cedette sotto i miei piedi e per un terribile secondo mi trovai a precipitare, finché con un tonfo doloroso atterrai in un'altra stanza. Questa era molto più buia e, anche dopo essermi abituato all'oscurità, non riuscivo a vedere che per pochi passi intorno a me. Per peggiorare la situazione, sentivo qualcosa muoversi nella stanza. Attanagliato dalla paura, inciampai e mi trovai di fronte a un'orrenda figura accovacciata, coperta da un'armatura e con in pugno una clava. Appena lo vidi sollevare la sua arma per colpirmi, un flusso di adrenalina mi schiarì le idee: con un unico fluido movimento misi una freccia sul mio arco, tesi la corda e scoccai la freccia. (Per mia fortuna, a scuola avevo seguito un corso di tiro con l'arco.) Con un sibilo e un urlo, il demonio (perché di un demonio si trattava) cadde a terra fremendo di rabbia. Con cautela uscii di lì, deciso a trovare una scala per tornarmene fuori dai Sotterranei del Destino. Pensavo alla mia casa accogliente e alla scrivania con l'articolo delle «(Ri)creazioni al calcolatore» da finire.

Mentre procedevo lentamente verso il punto in cui pensavo dovesse trovarsi una scala, colpii con lo stivale qualcosa che sembrava un mucchietto di pietre. Abbassai lo sguardo. Anche al buio qualcosa splendeva. Oro! Quasi quasi potevo esplorare un'altra stanza, una sola, prima di tornare in superficie...

A parte l'intensità dell'esperienza di immedesimazione, Avventuriero va al di là dei tradizionali giochi d'avventura sotto almeno due punti di vista. In primo luogo, la configurazione del terreno generata dal programma stesso e varia da una partita all'altra. In secondo luogo, Avventuriero fornisce al giocatore una

fino a quel punto, del livello dei sotterranei in cui il giocatore si trova (*si veda l'illustrazione di questa pagina*). Varie caratteristiche sono indicate da diversi caratteri della tastiera che appaiono sullo schermo. Per esempio, trattini e sbarrette rappresentano pareti, i segni «diesis» e i segni di percentuale rappresentano rispettivamente passaggi e scale, i più rappresentano entrate e i segni ∧ rappresentano trappole. Oggetti e caratteristiche della camera possono essere o non essere visibili a seconda che la camera sia illuminata o meno. Quando l'Avventuriero penetra in questo livello, le caratteristiche del livello non appaiono sullo schermo; perché vengano rappresentate, l'Avventuriero (rappresentato da un segno @) deve scoprirle. Così egli esplora, incappa in qualche trappola, perlustra parti di varie stanze e attraversa passaggi. Può anche incontrare la lettera *U*, che rappresenta un mostro chiamato il Cupo Hulk.

Sotto la mappa grafica, lo schermo mostra l'attuale situazione statistica dell'Avventuriero. Nella figura, l'Avventuriero si trova al venticinquesimo livello, ha accumulato 7730 pezzi d'oro, ha guadagnato 77 punti (ma gliene rimangono solo 25 a causa dei danni sofferti), ha accumulato forza fino al quindicesimo grado (18 è il massimo), indossa una armatura di classe 9 e ha 30 668 punti d'esperienza, sufficienti per porlo all'elevato tredicesimo ordine d'esperienza. Al venticinquesimo livello dei Sotterranei del Destino, all'Avventuriero non rimane che una scala da scendere per raggiungere il livello inferiore e cercare di impadronirsi dell'ambito Amuleto di Yendor. L'Avventuriero, però, deve prima sbarazzarsi del Cupo Hulk.

In questa sede possiamo analizzare solo superficialmente il gioco. Sarebbe possibile scrivere un intero libro di regole e consigli per giocare ad Avventuriero. Per ora, l'unico manuale del genere disponibile è un conciso opuscolo di otto pagine scritto da Michael C. Toy e Kenneth C. R. C. Arnold, *A Guide to the Dungeons of Doom (si veda la «Bibliografia» a pagina 106)*. Sfortunatamente, è disponibile solo per gli utenti del sistema a partizione di tempo VAX/UNIX. A quanto mi dice Arnold, è ora disponibile una versione di Avventuriero adatta per l'IBM PC e la si può richiedere (sotto la denominazione inglese di Rogue) alla A. I. Design, 201 San Antonio Circle, Suite 115, Mountain View, California 94040. Toy e Arnold, i creatori di Avventuriero, pensano evidentemente che il modo migliore per imparare il gioco sia giocarlo. È comunque possibile dare una descrizione a grandi linee delle caratteristiche generali del gioco.

A ogni livello, l'area di gioco è divisa in quadrati. L'Avventuriero, mentre esplora l'ambiente che lo circonda, occupa un quadrato alla volta. Il movimento all'interno delle stanze e dei passaggi è controllato mediante comandi formulati premendo alla tastiera lettere quali *h, j, k* e *l* che muovono l'Avventuriero di un unico quadrato in una delle quattro direzioni principali. Altri comandi producono un movimento diagonale o un movimento continuo in una data direzione. Per far salire o scendere una scala all'Avventuriero, il giocatore deve battere il carattere < o, rispettivamente, >.

Quando si scopre un oggetto vicino alla posizione dell'Avventuriero, per raccoglierlo automaticamente bisogna spostare l'Avventuriero sul quadrato in cui si trova l'oggetto. Se il giocatore vuole spostarsi su quel quadrato senza raccogliere l'oggetto, deve battere una *m* seguita dal carattere per la direzione appropriata. Battendo la *s*, si può anche far effettuare all'Avventuriero una ricerca delle trappole nei quadratini circostanti; con l'avvertenza, però, che la ricerca ha solo il 20 per cento di probabilità di portare effettivamente alla scoperta di una trappola.

Di tanto in tanto, l'Avventuriero deve fermarsi (battere un puntino) o mangiare (battere una *e*) per recuperare le forze spese nell'esplorazione o nella lotta contro i mostri. A parte una limitata quantità di cibo trasportata in uno zaino, l'Avventuriero non ha altro da mangiare che quello che può trovare per terra nei sotterranei (per quanto poco invitante possa sembrare).

Quando trova un pezzo di armatura, l'Avventuriero può prenderselo (mettendolo automaticamente nello zaino) o indossarlo (battere una *W*). Ovviamente, l'armatura dà all'Avventuriero maggiore protezione in combattimento, ma può anche essere stregata. Se l'armatura è stregata, ogni tentativo di togliersela (battere una *T*) è inutile senza una pergamena magica che rompe la maledizione. Tutti gli anelli magici trovati dall'Avventuriero possono essere indossati (battere una *P*) o tolti (battere una *R*) a meno che non siano stregati.

A volte l'Avventuriero è costretto ad abbandonare una fiaschetta e una pergamena o due (battere una *d* seguita da un carattere che rappresenta l'oggetto) perché lo zaino è pieno. (Un Avventuriero con lo zaino pieno non può raccogliere un'armatura.) Prima di abbandonare la fiaschetta, però, l'Avventuriero può berne l'ignoto contenuto sperando di trarne beneficio: alcune pozioni hanno un effetto curativo e altre mettono in grado l'Avventuriero di vedere il Mostro Invisibile. D'altro canto, la pozione può provocare uno stato di disorientamento; può capitare, ad esempio, che dopo aver dato all'Avventuriero il comando di muoversi verso nord, egli cominci a vagare in una direzione a caso. Allo stesso modo, può essere vantaggioso leggere una pergamena (battere una *r*) prima di gettarla: certe formule tolgono la maledizione da un'armatura.

Gli strumenti magici più impressionanti a disposizione dell'Avventuriero sono delle bacchette. A seconda del tipo di bacchetta raccolta, l'Avventuriero può sbarazzarsi di un mostro (battere una *z*) con vari effetti: può trasportarlo in un luogo a caso, lanciargli addosso globi infuocati o trasformarlo in un altro mostro. È meglio usare quest'ultimo tipo di bacchetta, detta bacchetta polimorfa, sui mostri più orribili: è di gran lunga preferibile trasformare il terribile Verme Purpureo in un pipistrello che viceversa.

Quando l'Avventuriero scopre un mostro, può essere opportuno lasciarlo stare: a volte un mostro sta dormendo e non attaccherà se lasciato in pace. Se però un combattimento appare inevitabile, si può impugnare un'arma (battere una *w*) e lottare all'ultimo sangue (battere una *f*), dopo aver spostato l'Avventuriero vicino al mostro. L'esito dello scontro dipenderà in modo probabilistico da fattori quali l'attuale livello di forza dell'Avventuriero, il suo grado di esperienza e la classe dell'armatura.

Tra i giocatori di Avventuriero, primeggia un giocatore non umano di grosso calibro: un programma per calcolatore che nel corso degli ultimi quattro anni si è dedicato a questo gioco rivaleggiando in prodezza con i migliori concorrenti alla ricerca dell'amuleto e dell'oro. Questo programma fornisce anche un'interessante opportunità di osservare un sistema esperto al lavoro.

Il 16 febbraio del 1984, all'Università del Texas di Austin, un Avventuriero controllato da un programma fece fuori tutti i mostri, ammassò un considerevole mucchio d'oro e ritornò con l'amuleto. Questo programma, dal nome ben poco magico di ROG-O-MATIC, diresse ogni passo dell'Avventuriero, ogni sua pausa, ogni lancio e ogni colpo.

ROG-O-MAGIC è la creazione di quattro studenti del dipartimento di scienze del calcolatore della Carnegie-Mellon University di Pittsburgh: Andrew Appel, Leonard Hamey, Guy Jacobson e Michael Mauldin. ROG-O-MATIC unisce fonti di conoscenza programmate a sistemi esperti per prendere decisioni sul da farsi in ogni immaginabile situazione sotterranea.

Quando un essere umano gioca ad Avventuriero, un flusso di comandi passa dalla tastiera al programma ROGUE grazie al sistema operativo. Il programma decide automaticamente quale mostro schierare contro l'Avventuriero, che configurazione usare per il prossimo livello dei sotterranei, e così via. Il programma trasmette queste informazioni, sempre attraverso il sistema operativo, allo schermo per tenere informato l'essere umano della situazione attuale.

Sostituendo un giocatore umano, il programma ROG-O-MATIC intercetta il flusso di caratteri da tastiera e trasmette per proprio conto caratteri al programma ROGUE. Quest'ultimo non ha nessuna idea (per così dire) del fatto che sta giocando un altro programma. Allo stes-

```
          --+-------                   -------------------------
          ...@.......|                |.......................|
          ...U.......|                |....^..................|
|......   ...........+#######+...              ..........     |
|..........................|                             .....|
|..........................|                              ...
 ---------------------+-                                 -+-
                                                           #
                              ####################
                              #
                            --+------
                          |.........|
                          |.........|
                          |.........|
                          |.......
                             ....
                             ----
```

Il venticinquesimo livello dei Sotterranei del Destino

@	l'Avventuriero	h	spostati di uno spazio a ovest
–|	pareti di stanze	j	spostati di uno spazio a sud
+	vano di porta	k	sposati di uno spazio a nord
#	passaggio tra stanze	l	spostati di uno spazio a est
·	pavimento di stanza	>	discendi la scala
%	scala	<	sali la scala
∧	trappola	m	spostati su qualcosa
)	arma	s	ricerca se ci sono trappole
]	pezzo di armatura	∧	identifica la trappola
*	oro	. /e	riposa/mangia
!	fiaschetta di pozione	W/T	indossa/togli l'armatura
?	pergamena magica	P/R	metti/togli l'anello
:	cibo	d	abbandona un oggetto
/	bacchetta magica	q	bevi una pozione
=	anello magico	r	leggi una pergamena
A–Z	le lettere maiuscole indicano abitanti dei sotterranei	z	elimina con la bacchetta magica
		t	getta un oggetto
		w	impugna un'arma
		f	combatti all'ultimo sangue

è

Simboli della mappa* (a sinistra) *e comandi per il gioco Avventuriero

vista in pianta del tratto, esplorato

so modo, anche l'informazione diretta dal programma ROGUE verso lo schermo viene diretta verso il programma ROG-O-MATIC, che può così avere la propria mappa dei Sotterranei del Destino.

Il programma ROG-O-MATIC è formato da 12 000 righe scritte nel linguaggio di programmazione C ed è ancora più lungo e complicato del programma ROGUE. ROG-O-MATIC iniziò nel 1981 con qualcosa che i suoi primi creatori, Appel e Jacobson, pensavano dovesse essere un «progetto semplice». Poco dopo che Mauldin si fu unito al gruppo, il primo programma che giocava ad Avventuriero passò attraverso una larga serie di modifiche dando luogo a diverse varianti, ciascuna delle quali aggiungeva qualche miglioramento nella tattica o nella strategia. Quando anche il quarto membro, Hamey, iniziò a contribuire allo sviluppo del programma, gli autori cominciavano a rendersi conto di aver creato, in effetti, un sistema esperto. Questi sistemi costituiscono un'applicazione chiave della cosiddetta Quinta Generazione di calcolatori, che secondo i progetti dovrebbe dare risultati commerciali alla fine degli anni ottanta. Un sistema esperto è progettato per incorporare e proiettare l'abilità umana in una grande varietà di aree, dalla medicina all'ingegneria.

Usando il tipo di architettura software adatta per sistemi esperti, i creatori di ROG-O-MATIC furono in grado di delineare e modificare il loro programma con relativa facilità. In particolare, organizzarono i vari tipi di conoscenza e abilità richiesti dall'Avventuriero in una gerarchia di sottosistemi diversi *(si veda l'illustrazione qui sotto)*.

Per esempio, un esperto di alto livello (chiamato Mischia) controlla la lotta durante il combattimento e un altro esperto di alto livello (chiamato Bersaglio) dirige la caccia dell'Avventuriero ai mostri. Entrambi questi esperti utilizzano un esperto di livello inferiore, detto *battaglia,* che effettua particolari attacchi oppure inizia una ritirata a seconda della situazione. L'esperto in battaglia chiama a volte in aiuto l'esperto in ritirata e quest'ultimo attinge invariabilmente a una fonte di conoscenza chiamata *perc,* un particolare algoritmo che studia il terreno alla ricerca del più breve percorso possibile per qualsiasi posizione specificata. In generale, gli algoritmi per il percorso più breve hanno avuto un notevole sviluppo e in questa particolare applicazione è accettabile solo l'algoritmo più veloce possibile: viene utilizzato quasi in continuazione nel corso dell'esplorazione dei Sotterranei del Destino da parte di ROG-O-MATIC. Le conoscenze sul terreno usate da *perc* vengono da *termap,* una struttura di dati che registra le caratteristiche del terreno scoperte fino a quel momento dall'Avventuriero nel corso della sua esplorazione di un particolare livello dei sotterranei. Infine, *termap* ricava tutte le sue conoscenze da *senso,* una struttura di dati di basso livello che contiene tutte le informazioni rilevanti fornite in uscita dal programma ROGUE.

Prima di elencare i compiti di altri esperti e di altre fonti di conoscenza, vorrei riesaminare più in dettaglio uno degli esperti, *battaglia.* Una volta che una battaglia è in corso, l'esperto in mischia richiede all'esperto in battaglia di decidere se attaccare o ritirarsi. Dato che il meglio del valore sta nella prudenza, l'esperto in battaglia stabilisce per prima cosa l'auspicabilità e fattibilità di una ritirata. Per farlo deve verificare l'esistenza di alcune condizioni preliminari:

1. L'Avventuriero non è attualmente sotto l'influenza della pozione che produce confusione (che potrebbe far partire direttamente l'Avventuriero verso il mostro).
2. L'Avventuriero non è già tenuto saldamente da un mostro.
3. Sarebbe possibile morire nel corso di una mischia (in questo caso è fortemente auspicabile evitare il conflitto).
4. L'esperto in ritirata può trovare una via di fuga (una ritirata è possibile).

Se valgono tutte e quattro le condizioni, ROG-O-MATIC affiderà l'incarico della ritirata dell'Avventuriero all'esperto in ritirata. Se invece non valgono, l'esperto in battaglia passa in rassegna una serie di possibilità aggressive.

1. Se è possibile che l'Avventuriero muoia nel corso di una mischia, se il mostro è vicino e visibile e se l'Avventuriero si trova ad avere una bacchetta per il teletrasporto, allora punta la bacchetta contro il mostro.
2. Se è possibile che l'Avventuriero muoia nel corso di una mischia, se il mostro è vicino e se l'Avventuriero ha una pergamena per il teletrasporto, allora legge la formula magica riportata sulla pergamena.

Se non vale né l'una né l'altra di queste condizioni, non rimane altra alternativa che andare almeno una volta allo scontro con il mostro, quindi il programma ROG-O-MATIC impegna arditamente in battaglia il suo Avventuriero. Forse è abbastanza forte per sopravvivere a uno scontro; in caso contrario, può anche rimanere ucciso.

Tra gli altri esperti usati dal programma c'è un esperto in esplorazione *(esplora)* che decide dove condurre la prossima esplorazione e che movimento usare. Ci sono poi un esperto in missili, che controlla il lancio di frecce, massi, lance e così via contro i mostri, un esperto nella scelta degli oggetti da raccogliere, chiamato *oggetto,* un esperto nella scelta dell'armatura da indossare, un esperto nell'uso della magia e un esperto, chiamato *salute,* che decide quando mangiare o riposarsi. *Oggmap* è una struttura di dati che registra la posizione e la storia di tutti gli oggetti incontrati fino a un certo punto, *invent* è un inventario di ciò che è contenuto nello zaino dell'Avventuriero e *intern* è un riconoscitore di stato interno che controlla se l'Avventuriero è pronto a uno sforzo.

Non c'è da sorprendersi se ROG-O-MATIC può giocare rapidamente una partita di Avventuriero. Dopo pochi minuti di tempo del calcolatore è tutto finito in un modo o nell'altro. Fino a oggi, solo alla Carnegie-Mellon il programma ha giocato più di 12 000 partite di Avventuriero; i dati statistici riportati dai quattro creatori di ROG-O-MATIC tendono a confortare l'affermazione che attualmente il programma supera in abilità la grande maggioranza degli esperti umani di Avventuriero. Per esempio, in un test effettuato alla Carnegie-Mellon durante un periodo di tre settimane nel 1983, ROG-O-MATIC ottenne un punteggio medio superiore a quello dei 15 migliori giocatori di Avventuriero dell'università. ROG-O-MATIC supera la maggior parte degli esseri umani e mostra, rispetto a essi, alcune differenze notevoli nello stile di gioco. Secondo Mauldin, il programma è prudente e privo di immaginazione; esplora in modo efficiente ed evita tutte le ricerche ridondanti che gli esseri umani probabilmente intraprendono. Il suo stile di lotta, invece, è piuttosto metodico e, se non è fortunato, non è in grado di imitare le ardite possibilità del gioco umano.

Sottosistemi esperti e fonti di conoscenza usati dal programma ROGUE

Consideriamo, per esempio, il fattore fortuna nella sua storica impresa del febbraio 1984 all'Università del Texas di Austin. Dopo aver superato gli incredibili pericoli dei Sotterranei del Destino, l'Avventuriero (naturalmente sotto il controllo di ROG-O-MATIC) trovò l'Amuleto di Yendor in un passaggio del ventiseiesimo livello. Tornando verso la superficie con il suo trofeo, l'Avventuriero incontrò al ventiduesimo livello uno dei peggiori mostri, un drago che sputa fuoco. L'Avventuriero estrasse la spada ma il drago soffiò una saetta di fuoco per primo. La fiamma mancò l'Avventuriero, che in quel momento si trovava nel vano di una porta, colpì una parete e rimbalzò direttamente contro il drago, sbruciacchiandolo gravemente. Con le forze che gli rimanevano, il drago avanzò verso l'Avventuriero, che lo finì con un colpo di spada e poi proseguì verso la superficie occupandosi lungo la strada di altri mostri meno pericolosi. Quando emerse alla luce del giorno, aveva in suo possesso non solo l'amuleto d'oro ma anche 6913 pezzi d'oro e altri oggetti.

Il lettore che, incuriosito dal precedente resoconto, voglia provare il divertimento offerto dal gioco può impegnarsi nel seguente rompicapo costruito appositamente da Mauldin. Oltre a essere un'introduzione ai piaceri di Avventuriero, illustra la difficoltà di inserire intelligenza nel programma ROG-O-MATIC.

Una stanza del ventiseiesimo livello dei Sotterranei del Destino è attualmente occupata dal Verme Purpureo (*P*), dal Grifone (*G*) e dall'Avventuriero (@). L'Amuleto di Yendor (,) si trova sul lato opposto di una scala (%) rispetto all'Avventuriero.

L'obiettivo dell'Avventuriero è afferrare l'amuleto e correre sulla scala senza rimanere ucciso. All'Avventuriero rimane un solo punto da usare in lotta e non può rischiare uno scontro con il Verme Purpureo o con il Grifone. Per fortuna, il Verme Purpureo è profondamente addormentato e l'Avventuriero, indossando l'Anello della Segretezza, deve solo evitare di calpestarlo per non svegliarlo. Il Grifone, d'altra parte, è quanto mai sveglio e in caccia dell'Avventuriero. Il Grifone non disturberebbe mai un mostro amico, ma vorrebbe molto occupare la stessa posizione dell'Avventuriero, che smembrerebbe con un solo colpo dei suoi crudeli artigli.

Tocca all'Avventuriero muovere. Ogni mossa dell'Avventuriero o del Grifone porta all'occupazione di uno degli otto punti adiacenti. Il Grifone è così ristretto di vedute da scegliere sempre una mossa lungo il percorso più diretto verso la sua preda. Se l'Avventuriero si sposta sulla scala è al sicuro dal Grifone e dal Verme Purpureo: le regole di cooperazione tra i mostri proibiscono l'inseguimento su altri livelli.

L'articolo del gennaio scorso sulla Torre di Hanoi e gli anelli cinesi ha provocato una grande varietà di risposte, che vanno dalla matematica alla metafisica. Parecchi lettori hanno presentato semplici soluzioni non ricorsive dei rompicapo. Tra essi Edward T. Price di Eugene, Oregon, il quale suggerisce di dipingere i dischi della Torre in due colori alternati, diciamo bianco e nero, secondo l'ordine di grandezza crescente. Il rompicapo si risolve allora facilmente aggiungendo la seguente regola a quelle originali: non mettere mai un disco su un disco dello stesso colore. Il rompicapo corrispondente degli anelli cinesi è stato risolto da Morris S. Samberg di Howard Beach, New York. Si alterna il movimento del primo anello (mettendolo o togliendolo) con il movimento di qualche altro anello. Ogni volta, solo un altro anello può essere mosso. Se il numero degli anelli è dispari, si parte con il primo anello; in caso contrario si parte con l'altro. Così Rob Hardy di Dayton, Ohio, riassume questa semplice soluzione:

Una soluzione iterativa
Non è più certo un divertimento:
Alternate il cambio dell'ultimo anello
Mettendone un altro in movimento.

Nessun lettore è stato in grado di trovare un semplice schema matematico dietro l'ordinamento del re Wen per gli esagrammi dell'*I Ching.* Molti lettori hanno osservato che gli esagrammi sono accoppiati attraverso le colonne in modo semplice, ma questo non spiega l'ordine globale. Homer E. Brown, un consulente di ingegneria elettromeccanica di Cary, North Carolina, ha prodotto un'analisi abbastanza suggestiva: a partire da ogni esagramma contare verso l'alto o verso il basso 10 esagrammi, passando dalla fine di una colonna all'inizio della successiva, se necessario. L'esagramma a cui si arriva in questo modo ha sempre una relazione semplice con quello di partenza. La regola, però, non è deterministica dato che non si sa mai se contare verso l'alto o verso il basso.

A dispetto di questa apparente relazione (che ha comunque un gran numero di eccezioni), io sono incline a ritenere che la configurazione del re Wen segua principi metafisici. Il nostro tentativo di proiettare le attuali preoccupazioni riguardo la scienza e la tecnologia sulle culture passate dà luogo a una visione distorta di ciò che era un tempo. Come regola generale, i sistemi «scientifici» di una volta svolgevano un ruolo strettamente ancillare rispetto a una visione del mondo fortemente religiosa. Opinioni analoghe sono proposte da Bernard X. Bovasso di Saugerties, New York. Lo *I Ching* si preoccupa dell'ordine del tempo, sostiene Bovasso, e così era per il re Wen. David White, un filosofo del Macalester College di St. Paul, Minnesota, ha fatto riferimento a una traduzione inglese dell'*I Ching,* opera di James Legge *(si veda la «Bibliografia» a pagina 106).* In un'appendice di quest'opera si traccia una visione metafisica che governerebbe l'ordine di tutti i 64 esagrammi. Per esempio, i primi tre sono collegati rispettivamente ai concetti di cielo, terra e caos. L'esagramma del caos, che denota ciò che viene considerato il disordine di tutte le cose create prese collettivamente, segue il cielo e la terra dato che tutte le cose create riempiono lo spazio tra il cielo e la terra.

Nel tentativo di risolvere il problema 9 per 9 della caccia, proposto nell'articolo di febbraio su squali e altri pesci, i lettori si sono forse trovati invischiati in complicazioni non necessarie. La griglia oceano che contiene i due squali e gli altri tre pesci dovrebbe essere di 9 per 10, quindi va aggiunta alla griglia un'altra linea verticale.